JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 5 MAY, 1951

In This Issue

	Page
Applications of the Radiation from Fast Electron Beams — H. Motz	527
Electrical Breakdown over Insulators in High Vacuum — Paul H. Gleichauf	535
Electrical Method for Investigating the Nature and Behavior of Small, Airborne, Charged Particles — J. H. Daniel and F. S. Brackett	542
Some Quantitative Data Bearing on the Scabbing of Metals under Explosive Attack — John S. Rinehart	555
Permanent Magnet Lenses — John H. Reisner	561
Internal Friction of Ferromagnetic Materials — A. Ferro and G. Montalenti	565
Electric Tunnel Effect across Thin Insulator Films in Contacts — Ragnar Holm	569
On the Diffraction of a Plane Electromagnetic Wave by a Paraboloid of Revolution — C. W. Horton and F. C. Karal, Jr.	575
Problem in Heat Conduction — Boris Podolsky	581
On the Accuracy of the Long-Range Ballistic Rocket — W. E. Frye	585
Probe Technique for the Measurement of Electron Temperature — M. A. Easley	590
Some Mechanical Properties of Graphite at Elevated Temperatures — C. Malmstrom, R. Keen, and L. Green, Jr.	593
Electromagnetic Scattering from Spheres with Sizes Comparable to the Wavelength — A. L. Aden	601
Electrode Reactions in the Glow Discharge — F. E. Haworth	606
Reflecting Surface to Simulate an Infinite Conducting Plane — S. J. Raff	610
Negative Coaxial Cylindrical Corona Discharges in Pure N_2, O_2, and Mixtures Thereof — Charles G. Miller and Leonard B. Loeb	614
Plastic Deformation of Pure Single Crystals of Lead and Copper — Peter W. Neurath and James S. Koehler	621
Relation between Gas Evolution and Physical Properties of Liquids — V. G. Szebehely	627
Electric and Magnetic Constants of Metallic Delay Media Containing Obstacles of Arbitrary Shape and Thickness — Seymour B. Cohn	628
Lattice and Grain Boundary Self-Diffusion in Silver — R. E. Hoffman and D. Turnbull	634
On Shock-Wave Phenomena: Waterlike Substances — R. J. Seeger and H. Polachek	640
Scattering Phenomena in Electron Microscope Image Formation — C. E. Hall	655
X-Ray Measurement of Long-Range Order in β-AgZn — L. Muldawer	663
New X-Ray Diffraction Method for Studying Imperfections of Crystal Structure in Polycrystalline Specimens — Alfred J. Reis, Jerome J. Slade, Jr., and Sigmund Weissmann	665
Attenuation of Radio Signals Caused by Scattering — A. H. LaGrone, W. H. Benson, Jr., and A. W. Straiton	672
On the Diffraction of Electromagnetic Waves by Two Conducting Parallel Half-Planes — M. G. Cheney, Jr. and Robert B. Watson	675
Letters to the Editor:	
Radiofrequency Mass Spectrometer — Paul Schissel	680
Oriented Crystals in the Rolled Plates of Alloys — Shigeto Yamaguchi	680
Studies of Thin Films by Electron Diffraction — N. R. Mukherjee and Oliver Row	681
Comment on Nomenclature in "Electron Microradiography of Electrodeposited Metals" — G. M. Corney	682
Effect of Auxiliary Current on Transistor Operation — H. J. Reich, P. M. Schultheiss, J. G. Skalnik, T. Flynn, and J. E. Gibson	682
Books	682
Other Books Received	688
Announcements	
New Appointment	554
Awards and Honors	560
Summer Program in Nuclear Engineering at North Carolina State College	574
New Booklets	609
Summer Program at Oak Ridge Institute of Nuclear Studies	654
Publication Charge for Letters to the Editor	679
Index to Advertisers	xxx

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Paul Kirkpatrick
F. W. Loomis
J. R. Oppenheimer
George B. Pegram
I. I. Rabi
Duane Roller
Harold K. Schilling
Frederick Seitz
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 5 May, 1951

Applications of the Radiation from Fast Electron Beams

H. Motz
Microwave Laboratory, Stanford University, California
(Received July 3, 1950)

The radiation from fast electron beams passing through a succession of electric or magnetic fields of alternating polarity is examined. The radiation of maximum frequency is emitted in the forward direction. If the deflecting fields are not too large, a semiqualitative argument shows that the maximum frequency is the lowest possible harmonic. The frequencies emitted are determined by studying the Doppler effect, and the angular distribution of radiated energy as well as the total radiation are calculated in a simple straightforward manner with reference to well-known formulas of special relativity. The question of the coherence of the radiation is discussed. The spectral distribution of radiated energy is then calculated more exactly. It is concluded that several applications of the radiation appear possible. A scheme for obtaining millimeter-waves of considerable power is outlined. The upper limit of the power in a band extending down to a wavelength of 1 millimeter is calculated to be of the order of several kilowatts for a beam of one ampere and an energy of 1.5 megavolt. The use of the radiation for speed monitoring of beams with energies up to 1000 megavolt is discussed.

I. INTRODUCTION

AN electron, traveling through a stationary electric or magnetic field distribution, radiates. The frequency spectrum of the emitted radiation depends on the speed of the electron. We shall see that the entire spectrum of electromagnetic radiation starting from microwaves and extending to hard x-rays may be easily obtained from electrons with speeds ranging from a megavolt, say, to 1000 Mev.

Several applications suggest themselves. We shall discuss them under the following headings:

(1) production of energy in rather inaccessible spectral bands, *viz.*, millimeter to infrared radiation;*

(2) speed monitoring for electron beams produced by linear or other accelerating devices;

(3) speed measurement for fast individual electrons or other particles (mesons, protons).

We shall also briefly consider the radiation emitted by an electron moving through an electromagnetic wave traveling in the opposite direction. This occurs, e.g., in a linear accelerator in which a wave is reflected from the output end and travels back to the starting section.

The stationary field distribution we have in mind is a succession of electric or magnetic fields of alternating polarity, regularly spaced, as in Fig. 1. The radiation in transverse fields turns out to be much larger than that emitted by electrons crossing an array of longitudinal fields. Hence, we shall be mainly interested in the case of transverse fields, as shown in Fig. 1. The spatial field distribution may be fourier analyzed. Let the fundamental wavelength be l_0. It will be shown that the frequency of the emitted radiation varies with the angle of observation. The fundamental component of the radiation emitted under the angle θ (see Fig. 1) has a frequency

$$\omega=\Omega_\theta=2\pi\beta c/l_0(1-\beta\cos\theta). \quad (1)$$

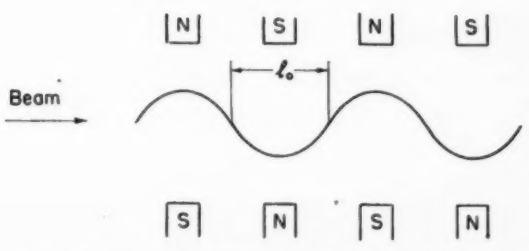


FIG. 1. Schematic arrangement of undulator magnets.

* *Notes added in proof:* According to reports of the Electronics Research Laboratory at M.I.T., P. D. Coleman is working on such a scheme. Prior reference to the problem treated in this paper has also been found in a paper by V. L. Ginsburg, Radiation of microwaves and their absorption in air, Bull. Acad. Sci. U.R.S.S. Ser. Phys. 9 (1947), No. 2, 165 (in Russian).

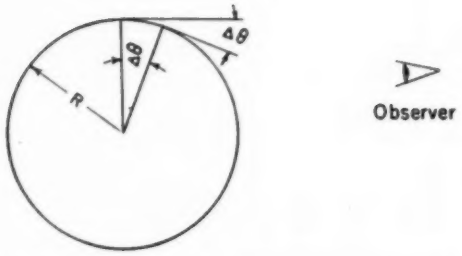


FIG. 2. Illustrates the radiation cone emanating from an electron in a circular orbit.

In the extreme relativistic case, when the electron velocity is near the velocity of light and the accelerations and decelerations suffered by the electrons in the periodic field are considerable, most of the energy is in a high harmonic of this fundamental frequency, which is determined by the maximum instantaneous acceleration or deceleration.[1] However, even for extremely relativistic electron speeds, the energy is predominantly in the fundamental provided that the deflecting fields are small, i.e., of the order of a megavolt/cm.

To show this we run through the following qualitative argument. Assume that the electrons in the periodic field describe arcs of circles, reminiscent of a sine wave. The radiation is mainly forward and contained in a narrow cone of opening

$$\Delta\theta = m_0c^2/E, \tag{2}$$

where E is the beam energy. As the electrons move along, the cone swings to and fro and may intermittently pass out of the view of an observer situated on the axis. This observer may then receive short pulses of radiation only, which explains why the energy is radiated into a higher harmonic.

Consider Fig. 2. As the electron moves along a circular arc of radius R, the radiation cone sweeps past the observer and is out of his view when the electron has moved through $\Delta\theta$ radians. Thus, the time during which radiation from the electron is received is given by

$$\tau = (1/c)R\Delta\theta(1-\beta) \sim \tfrac{1}{2}(1/c)(1-\beta^2)R\Delta\theta, \tag{3}$$

the radius of the circle is given by $m\beta^2c^2/R = e\mathcal{E}$, where $\mathcal{E}$ is the deflecting field. Hence,

$$\tau = (1-\beta^2)^{\frac{1}{2}}m\beta^2c^2/2ce\mathcal{E}. \tag{4}$$

The wavelength of the radiation is then approximately given by

$$\lambda_1 = \tfrac{1}{2}(1-\beta^2)^{\frac{1}{2}}\beta^2mc^2/e\mathcal{E} = \tfrac{1}{2}(1-\beta^2)\beta^2m_0c^2/e\mathcal{E}. \tag{5}$$

The fundamental frequency (1) corresponds to a wavelength

$$\lambda_2 = l_0(1-\beta)/\beta, \tag{6}$$

in the forward direction.

If the wavelength λ_1 is small compared with λ_2, the radiation is indeed mostly contained in a high harmonic. If, however,

$$\lambda_1 \geqq \lambda_2, \tag{7}$$

the radiation is mostly of fundamental frequency. This means we can expect to find most of the energy in the fundamental if

$$e\mathcal{E}l_0/m_0c^2 < 1. \tag{8}$$

In the following section we shall determine the radiated energy and frequencies under the assumption that Eq. (8) holds. This will first be done in a simple manner which clearly shows the physical principles involved. Later a more complete analysis of the frequency spectrum is carried out.

II. FREQUENCY OBSERVED IN THE LABORATORY SYSTEM

1. The Doppler Effect

To determine the observed frequencies we must apply the theory of the relativistic Doppler effect which we shall now review briefly. Let the moving system E have primed coordinates and let it emit a spherical wave of amplitude

$$S' = (A'/r')\exp[2\pi i\nu'(t'-r'/c)+i\delta']. \tag{9}$$

Let the observer belong to a system L (laboratory system) with unprimed coordinates z, y, r, l, and let the directions z and z' be parallel to the relative (constant) velocity of the two systems. We call this velocity v and define $\beta = v/c$.

Let the observer be located at the origin 0 of system L and let the direction $\langle 00'\rangle$ define an angle θ (in L) with the direction of movement, the positive z-direction. Let the corresponding angle between $\langle 00'\rangle$ and z be θ' in E. Then

$$r' = z'\cos\theta' + y'\cos\theta'$$

and

$$S' = (A'/r')\exp\{2\pi i\nu'[t'-(z'\cos\theta' + y'\sin\theta')/c]+i\delta\}'. \tag{10}$$

The observer in L interprets this as a spherical wave

$$S = (A/r)\exp\{2\pi i\nu[t-(z\cos\theta + y\sin\theta)/c]+i\delta\}. \tag{11}$$

The primed and unprimed coordinates are, of course, related by the Lorentz transformations

$$z' = (z-vt)/(1-\beta^2)^{\frac{1}{2}}, \quad t' = (t-vz/c^2)/(1-\beta^2)^{\frac{1}{2}}, \quad y = y', \quad x = x'. \tag{12}$$

Thus we can eliminate the primed coordinates in Eq. (10).

Comparing Eqs. (10) and (11) we must get relations which hold good for any time and place; thus, the coefficients of z, y, t must be equal. In this way one obtains

$$\nu = \nu'(1+\beta\cos\theta')/(1-\beta^2)^{\frac{1}{2}}, \tag{13}$$

$$\cos\theta = (\cos\theta'+\beta)/(1+\beta\cos\theta'), \tag{14a}$$

$$\nu = \nu'(1-\beta^2)^{\frac{1}{2}}/(1-\beta\cos\theta). \tag{14b}$$

[1] Compare L. I. Schiff, Rev. Sci. Instr. **17**, 6 (1946); J. Schwinger, Phys. Rev. **75**, 1912 (1949).

2. Frequency Seen by the Moving Electrons

(a) *By Riding through a Succession of Magnetic or Electric Fields of Alternating Polarity (See Fig. 1)*

The fundamental fourier component of the field has a wavelength l_0 in L, the length over which the periodic field extends is l and the number of full fundamental waves $n=l/l_0$.

An observer at rest in E (the system attached to the electron) takes a time $t'=l'/v$ to cover the distance l' as he rides past it. He receives $n=l/l_0=l'/l_0'$ fundamental waves; hence, he sees a frequency (number of waves/unit time)

$$\nu'=l'v/l_0'l'=v/l_0',$$

but

$$l_0'=(1-\beta^2)^{\frac{1}{2}}l_0.$$

Hence, he sees a frequency

$$\nu'=v/l_0(1-\beta^2)^{\frac{1}{2}}. \tag{15}$$

(b) *Situation in a Section of a Linear Accelerator Tube Electron Moving with Velocity v Relative to the L-System, Encountering a Traveling Wave with Velocity v Relative to the L-System*

Let there be ν^* waves emitted by the end of the accelerator tube (reflected waves) in unit L-time.

Measure coordinates and time in the system W attached to the wave and denote them by double primes.

During the time dt'' the wave has moved a distance

$$dz''=-\beta c dt''=-\beta c dt/(1-\beta^2)^{\frac{1}{2}}.$$

During the time dt'' the W-system has received ν^*dt full waves emitted by the accelerator end, the length of a full wave appears to be (for a W-observer)

$$\beta c/\nu^*(1-\beta^2)^{\frac{1}{2}}$$

The relative velocity between electron and wave, by Einstein's velocity addition theorem is

$$u/c=\{1-[(1-\beta^2)/(1+\beta^2)]^2\}^{\frac{1}{2}}. \tag{16}$$

Thus in time dt' the electron will cover a distance

$$dz_1''=-udt_1''=udt'/(1-u^2/c^2)^{\frac{1}{2}}$$

along the wave and so during time dt' the electron encounters

$$\nu'dt'=udt'(1-\beta^2)^{\frac{1}{2}}\nu^*/(1-u^2/c^2)^{\frac{1}{2}}\beta c$$

waves.

We find that this latter expression equals

$$\nu'dt'=2\nu^*dt'/(1-\beta^2)^{\frac{1}{2}}.$$

Thus, the frequency as seen by the electron is

$$\nu'=2\nu^*/(1-\beta^2)^{\frac{1}{2}}. \tag{17}$$

3. Frequency Received in L-System

In case (a): the frequency emitted by the electron is ν' in its own system. This is much higher than microwave frequency in the cases which interest us. Thus, we can treat it as a light wave propagated with velocity c and calculate the frequency ν received by the observer in L by means of the Doppler formula (13).

$$\nu=v(1+\beta\cos\theta')/l_0(1-\beta^2)=v/l_0(1-\beta\cos\theta). \tag{18}$$

In case (b): We apply the same Doppler formula (13) and obtain

$$\nu=2\nu^*(1+\beta\cos\theta')/(1-\beta^2)=2\nu^*/(1-\beta\cos\theta). \tag{19}$$

III. ENERGY RADIATED BY AN ELECTRON

Total energy calculated in the E-system and later referred to the L-system.—We want to calculate the energy radiated by an electron riding through a succession of electric or magnetic fields of alternating polarity. We shall call the arrangement an undulator.

First we shall calculate the energy radiated as it appears to an observer moving with the electron (E-system).

We apply the well-known formula for the Poynting vector

$$\mathbf{S}'=e^2v'^2\sin^2\psi'/4\pi c^3r'^2, \tag{20}$$

which measures the energy radiated per unit time through unit area of a spherical surface element at distance r' from the source whose normal makes an angle ψ' with the acceleration vector.

To obtain the total energy radiated we have to multiply Eq. (20) with the time l'/v during which the electron radiates and integrate Eq. (20) over the sphere. We obtain

$$W'=4\pi\tfrac{2}{3}(e^2\dot{v}^2/4\pi c^4)l'/\beta=\tfrac{2}{3}(e^2\dot{v}'^2/c^4)l'/\beta. \tag{21}$$

Note that we have used primed quantities because we have calculated the energy in the E-system.

The undulator may consist of a succession of magnetic fields of strength H with alternating polarity, let us call it the magnetic undulator. Alternatively, we may consider an analogous electric undulator—in this case let the maximum field strength be E. (More exactly, E and H are the amplitudes of the fundamental component of the spatial field distribution).

The fields which the electron sees are given by the relativistic field transformation formulas:

$$H_z'=H_z,\quad H_y'=(H_y+\beta E_z)/(1-\beta^2)^{\frac{1}{2}},\quad H_x'=(H_x-\beta E_y)/(1-\beta^2)^{\frac{1}{2}}. \tag{22a}$$

$$E_z'=E_z,\quad E_y'=(E_y-\beta H_x)/(1-\beta^2)^{\frac{1}{2}},\quad E_x'=(E_x+\beta H_y)/(1-\beta^2)^{\frac{1}{2}}. \tag{22b}$$

Looking at these expressions we see that the longitudinal field components are unchanged, while the transverse components increase with β.

Accordingly, the electron acceleration in E is much higher for transverse fields.

Therefore, we first consider an undulator with purely transverse field:

(a) electric,
(b) magnetic.

TABLE I.

Beam energy in Mev	Beam current	Power produced	Minimum fundamental wavelength
1000	1 μa	2.8 mW	25A
100	100 μa	2.8 mW	2500A
10	10 ma	2.8 mW	$2.5\ 10^{-3}$ cm

Consider case (a). The transverse field $E_y = E$. Hence,

$$E' = E/(1-\beta^2)^{\frac{1}{2}}, \tag{23}$$

the acceleration

$$\dot{v} = (e/m_0 E)/(1-\beta^2)^{\frac{1}{2}} \quad (m_0 \text{ rest mass}). \tag{24}$$

The length l appears shortened to the observer in E

$$l' = l(1-\beta^2)^{\frac{1}{2}}. \tag{25}$$

With these expressions for l' and v the energy radiated in E becomes

$$\tfrac{2}{3}e^2E^2l/m_0c^2(1-\beta^2)^{\frac{1}{2}}. \tag{26}$$

Thus, energy must now be transformed into the L-system by means of transformation formulas

$$W = (W'-\beta p_x')/(1-\beta^2)^{\frac{1}{2}},\ p_x = (p_x'-\beta W')/(1-\beta^2)^{\frac{1}{2}} \tag{27}$$

for the energy momentum tensor. The momentum in the E-system is zero. Thus,

$$W = (2/3\beta)(e^2/m_0c^2)^2E^2l/(1-\beta^2). \tag{28}$$

It is interesting to note that this is just the field energy swept out by the electron with the classical radius e^2/mc^2.

Now consider case (b). Let there be a magnetic field $H = H_z$. The field H_z' does not produce a force on the electron at rest in the E-system. But there is an electric force $E_y' = -\beta H_z/(1-\beta^2)^{\frac{1}{2}}$ which produces an acceleration

$$\dot{v} = -e\beta H/m_0(1-\beta^2)^{\frac{1}{2}},$$

and the energy W becomes

$$W = \tfrac{2}{3}(e^2/m_0c^2)^2\beta H^2l/(1-\beta^2). \tag{29}$$

IV. RADIATION BY THE ENTIRE BEAM: QUESTION OF COHERENCE

We show briefly, that for large β-radiation from N electrons is N times the radiation from the single electron, and not N^2 times this quantity, as one might hope to get from a coherently radiating beam. The field produced by N electrons is, of course, a superposition of the individual fields

$$E = E_1+E_2+E_3+E_4+\cdots,$$

and the energy is proportional to the square of this

$$(E_1 = E_2+E_3+\cdots)^2;$$

but the contributions of individual electrons are random in phase, and therefore the cross terms of the squared bracket drop out, and we are left with the sum of N squares

$$E_1^2+E_2^2+E_3^2+\cdots,$$

which is N times the energy radiated by the individual electron. This must be so, unless it be possible to bunch the electrons so that a whole bunch arrives at successive elements of the grating simultaneously. However, this is hard to realize. In the E-system the distance l_0 appears shortened to $l_0(1-\beta^2)^{\frac{1}{2}}$ and the bunches would have to be short compared with this to be effective.

It turns out that there is a small region of the spectrum which is interesting, while the bunching is yet possible: the millimeter wave range.

V. NUMERICAL VALUES

Let the fundamental wavelength of the undulator field be $l = 1$ cm. Let us further consider a magnetic undulator with $H = 15{,}000$ gauss. Then the total energy radiated incoherently amounts to the values shown in Table I.

For any harmonic the radiation of maximum frequency (minimum wavelength) is emitted in the forward direction ($\theta = 0$).

Another way of stating the result is to express the radiation per electrons in terms of quanta of maximum fundamental frequency. We can then say that the total radiated (emitted into the entire spectrum radiated) amount to 3 such quanta.

The field strength E of an electric undulator can probably not be made much greater than 1 Mev/cm, and under these conditions the energy output is about 1/25 of that of the magnetic undulator considered above.

VI. LONGITUDINAL FIELD ARRAYS

The longitudinal fields are the same in the E and L systems, and therefore the radiated energy is smaller by a factor $1-\beta^2$. This, no doubt, is the reason why the linear accelerator compares so favorably with other devices relying on circular electron orbits involving transverse fields. For the electric case we obtain an energy

$$W = \tfrac{2}{3}(e^2/m_0c^2)^2E^2l/\beta. \tag{30}$$

VII. THE ANGULAR DISTRIBUTION OF RADIATED ENERGY

The radiation of an electron in arbitrary motion is given[2] by

$$\frac{\mathbf{H}}{e} = \frac{(\mathbf{r}\times\dot{\mathbf{v}})(\mathbf{r}\cdot\mathbf{v}-cr)-(\mathbf{r}\times\mathbf{v})(\mathbf{r}\cdot\dot{\mathbf{v}})}{(cr-\mathbf{r}\cdot\mathbf{v})^3}, \tag{31a}$$

$$\frac{\mathbf{E}}{e} = \frac{(\mathbf{r}c-r\mathbf{v})(\mathbf{r}\cdot\dot{\mathbf{v}})-\dot{\mathbf{v}}\mathbf{r}(\mathbf{r}\cdot\mathbf{v}-cr)}{(cr-\mathbf{r}\cdot\mathbf{v})^3}. \tag{31b}$$

[2] See P. Frank and R. v. Mises, *Differential- und Integralgleichungen der Mechanik und Physik*, Vol. II, p. 788, formula 19. (Mary S. Rosenberg, New York 25, N. Y.)

Since

$$H=\mathbf{r}\times\mathbf{E}/r \quad \text{and} \quad \mathbf{r}\cdot\mathbf{E}=0,$$

it is seen immediately that E and H are both perpendicular to r and to each other. We assume v to be parallel to z and $\dot{v}$ parallel to y (this is the transverse field case). With definitions,

$$\begin{aligned} \mathbf{r}\times\mathbf{y}/y&=r\cos\psi, & \mathbf{r}\cdot\mathbf{z}/z&=r\cos\theta,\\ \mathbf{r}\times\mathbf{y}/y&=r\sin\psi, & \mathbf{r}\cdot\mathbf{z}/z&=r\sin\theta, \end{aligned} \tag{32}$$

we find for the Poynting vector (which has the dimension of power)

$$\mathbf{S}=\frac{c}{4\pi}(\mathbf{E}\times\mathbf{H})=\frac{\mathbf{r}e^2|\dot{v}|^2}{4\pi rc^3r^2}\left[\frac{1}{(1-\beta\cos\theta)^4}-\frac{(1-\beta^2)\cos\psi}{(1-\cos\theta)^6}\right]. \tag{33}$$

For the case of an electron moving in the direction of the acceleration vector (longitudinal undulator fields) it is found that

$$S=e^2\dot{v}^2\sin^2\theta/4\pi c^3r^2(1-\beta\cos\theta)^6. \tag{34}$$

To check the previous calculations carried out in the E-system we have to integrate Eq. (34) over the surface of the sphere after multiplication with the time during which the radiation is received in order to find the total energy. This time depends on the angle of observation. Consider formula (5a) for the Doppler effect of frequency. The frequency received varies with the angle θ, between the observer and the z-axis. However, the electron encounters n full waves of the grating and the observer cannot receive more or less oscillations of fundamental frequency from the radiating electron. Thus, the time during which radiation is received by the observer varies with θ. It is given by

$$T_\theta=l(1-\beta\cos\theta)/\beta c. \tag{35}$$

This is easily seen from Fig. 3 where the arrow points to an observer at large distance. It is seen that the radiation coming from β is delayed by a time interval $-(l/c\beta)\cos\theta$.

For the integration over the angles it is preferable to introduce an azimuth χ in the y-x plane (Fig. 4). From the cosine theorem of solid geometry

$$\cos\psi=\cos\chi\sin\theta. \tag{36}$$

The total energy radiated by the transverse undulator is thus given by

$$\frac{le^2\dot{v}^2}{4c^4\beta}\left[\int_{-1}^{+1}\frac{2du}{(1-\beta u)^3}-(1-\beta^2)\int_{-1}^{+1}\frac{(1-u^2)du}{(1-\beta u)^5}\right], \tag{37}$$

which indeed yields

$$\tfrac{2}{3}e^2\dot{v}^2l/c^4\beta(1-\beta^2)^2. \tag{38}$$

The corresponding total energy radiated in a longitudinal field array turns out to be

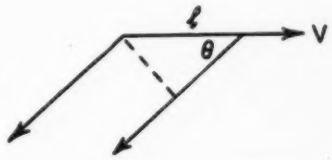


FIG. 3. Illustrates retardation of radiation from a moving electron.

$$\tfrac{2}{3}e^2\dot{v}^2l/c^4\beta(1-\beta^2)^3. \tag{39}$$

Of course, in the transverse case (38) we have to use the transverse mass

$$m_0/(1-\beta^2)^{\frac{1}{2}}, \tag{40}$$

while in the longitudinal case (39) we have to use the longitudinal mass

$$m_0/(1-\beta^2)^{\frac{3}{2}} \tag{41}$$

when computing the acceleration and thus fall back on formulas (28) and (29), respectively, and on Eq. (30).

VIII. CALCULATION OF THE FREQUENCY SPECTRUM RADIATED BY THE UNDULATOR

1. A Fundamental Radiation Formula

In order to calculate the frequency spectrum in more detail, two different procedures are available. We choose to work with a fourier expansion of the current density rather than with fourier expansions of the fields. In this case a convenient starting point is an expression for the average energy $P_{\mathbf{k}\omega}(\mathbf{r}')d\omega$ emitted into a frequency range between ω and $\omega+d\omega$ by a current with density distribution $J(\mathbf{r}')$. Let the observer be situated at a point with radius vector $\mathbf{r}$ drawn from the origin of the coordinate system. Let the distance $|\mathbf{r}'-\mathbf{r}|$ be large compared with the extension of the radiating current system, and let $\mathbf{k}$ be a vector of absolute value $|k|=\omega/c$ which points from the current element at $\mathbf{r}'$ to the observer at $\mathbf{r}$. Then the average energy is given by

$$P_{\mathbf{k}\omega}(\mathbf{r})d\omega=(k^2/2\pi r^2c)\left|\int J_{\perp\mathbf{k}\omega}(\mathbf{r}')e^{-i\mathbf{k}\cdot\mathbf{r}'}d\tau'\right|^2 d\omega, \tag{42}$$

where $J_{\perp\mathbf{k}}(\mathbf{r}')$ is the fourier amplitude of the current density vector projected on to the plane perpendicular to $\mathbf{k}$. For a derivation of this formula, see L. I. Schiff, *Quantum Mechanics*.[3]

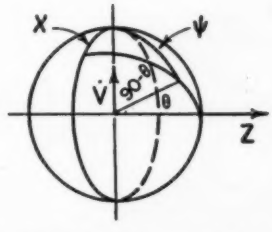


FIG. 4. Diagram showing relation between the spherical angles.

[3] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), p. 264.

2. Electron Orbits in an Undulator

To calculate the radiation emitted by fast electrons crossing the undulator we first determine the motion of the electrons in a field given by the fundamental components of a fourier expansion of the undulator field. Let this be an electric field,

$$E_x=\mathcal{E}\sin(2\pi/l_0)z=\mathcal{E}\sin(2\pi\beta c/l_0)t=\mathcal{E}\sin\omega_0 t,\qquad \omega_0=2\pi\beta c/l_0, \quad (43)$$

in the transverse direction x. The exact equations for the relativistic motion,

$$d(mv_x)/dt=e\mathcal{E}\sin(2\pi/l_0)z,\quad v_x=dz/dt;$$
$$d(mv_z)/dt=0,\quad m=m_0/[1-(v_x^2+v_z^2)/c^2]^{\frac{1}{2}},\quad v_z=dx/dt,$$

are difficult to solve. We shall, in fact, assume that the transverse velocity is small compared with the longitudinal one, and also that the change in longitudinal velocity is negligible. We then find that the transverse velocity v_x is given by

$$v_x=-e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}\cos(\omega_0 t)/m_0\omega_0, \quad (44)$$

and the transverse deflection x by

$$x=e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}\sin(\omega_0 t)/m_0\omega_0^2=-A\sin\omega_0 t. \quad (45)$$

A better approximation is afforded by the assumption that the absolute value of the velocity is constant along the path. With this constant velocity $v=\beta c$ we have again, putting at first $z=vt$,

$$v_x=e(1-\beta^2)^{\frac{1}{2}}\cos(\omega_0 t)/m_0\omega_0$$

and for the longitudinal velocity we obtain

$$v_z=v(1-v_x^2/v^2)^{\frac{1}{2}}=\beta c(1-\tfrac{1}{2}v_x^2/v_2)$$
$$=v[1-(1-\beta^2)^{\frac{1}{2}}e^2\mathcal{E}^2\cos^2(\omega_0 t)/2\beta^2c^2\omega_0^2m_0^2]. \quad (46)$$

Integrating this with respect to t we find

$$z=(v-a^2)t-a^2\sin(2\omega_0 t)/2\omega_0,$$
$$a^2=\tfrac{1}{4}(1-\beta^2)e^2\mathcal{E}^2/\beta^2c^2m_0^2\omega_0^2\ll\beta^2. \quad (47)$$

We now assume t in the form

$$t=(z/v)-\alpha f(z), \quad (48)$$

where α is a small quantity and $f(z)$ is an unknown function which we now set out to determine. Substituting Eq. (48) into Eq. (47) we find

$$(za/v)f(v-a^2)\alpha f(z)$$
$$+(a^2/2\omega_0)\sin2\omega_0[(z/v)-\alpha f(z)]=0. \quad (49)$$

From this we find by expanding the sine function in terms of the small quantity α,

$$-\alpha f(z)=\frac{1}{v-a^2}\left\{\frac{za^2}{v}+\frac{a^2}{2\omega_0}\sin\frac{2\omega_0 z}{v}\right\}\left\{1+\frac{a^2}{v}\cos2\omega_0\frac{z}{v}\right\}, \quad (50)$$

the leading term of this expression is of order $a^2/2\omega_0 v$. We observe that we should now recalculate v_x using Eq. (57). It is easily seen that this would not change the terms of order $a^2/2\omega_0 v$ in Eq. (50).

We shall proceed to calculate the radiation neglecting terms of this order. It will then be possible to see what the corrections amount to. However, we know already that longitudinal oscillations of the electrons give rise to much weaker radiation than transverse ones.

3. Calculation of the Radiation

The components of current density J_x and J_z due to a single electron may be given in the form of products of δ-functions,

$$J_x=ev_x\delta(y)\delta(x+A\sin\omega_0 z/\beta c)\delta(z-c\beta t), \quad (51)$$

$$J_z=e\beta c\delta(y)\delta(x+A\sin\omega_0 z/\beta c)\delta(z-\beta c t). \quad (52)$$

The fourier components of current density of angular frequency ω, i.e., the amplitudes $J_{x\omega}$, $J_{z\omega}$, of the fourier integral expansions

$$J_z(\mathbf{r},t)=\int_0^\infty(J_{z\omega}(\mathbf{r})e^{i\omega t}+\text{c.c.})d\omega, \quad (53)$$

$$J_x(\mathbf{r},t)=\int_0^\infty(J_{x\omega}(\mathbf{r})e^{i\omega t}+\text{c.c.})d\omega, \quad (54)$$

of Eqs. (51), (52) are given by

$$J_{x\omega}=(ev_x/2\pi\beta c)\delta(y)\delta(x+A\sin\omega_0 z/\beta c)e^{i\omega z/\beta c}, \quad (55)$$

$$J_{z\omega}=e[\delta(y)/2\pi]\delta(x+A\sin\omega_0 z/\beta c)e^{i\omega z/\beta c}. \quad (56)$$

This can be seen by substituting Eqs. (55), (56) back into Eqs. (53), (54) and using well-known rules concerning the operation with δ-functions.

Let the direction of observation $\mathbf{k}$ have polar angles ϕ, θ, with respect to the z-axis. $J_{\perp\mathbf{k}\omega}$ is given by

$$J_{\perp\mathbf{k}\omega}=J_{x\omega}(1-\cos^2\phi\sin^2\theta)^{\frac{1}{2}}\mathbf{u}_1+J_{z\omega}\sin\theta\mathbf{u}_2, \quad (57)$$

where $\mathbf{u}_1$ and $\mathbf{u}_2$ are unit vectors in the directions of the components $\perp\mathbf{k}$ of $J_{x\omega}$ and $J_{z\omega}$.

We substitute Eq. (57) into Eq. (42) and find

$$P_{\mathbf{k}\omega}=\frac{k^2e^2}{8\pi^3r^2c}\left|\int\int\delta\left(x'+A\sin\frac{\omega_0 z'}{\beta c}\right)\right.$$
$$\exp[-ik(x'\sin\theta\cos\phi+z'\cos\theta)+i\omega z'/\beta c]$$
$$\times[\mathbf{u}_1\sin\theta+\mathbf{u}_2(v_x/\beta c)(1-\cos^2\phi\sin^2\theta)^{\frac{1}{2}}]$$
$$\left.\times dx'dz'\right|^2, \quad (58)$$

where the integration over dy', which is trivial, has been carried out. We have further

$$P_{\mathbf{k}\omega}=\frac{k^2e^2}{8\pi^3r^2c}\left|\int\left(\exp\left\{-ik\left[-A\sin\frac{\omega_0 z'}{\beta c}\sin\theta\cos\phi\right.\right.\right.$$
$$\left.\left.\left.-z'\left(\frac{1}{\beta}-\cos\theta\right)\right]\right\}\right)[\mathbf{u}_1\sin\theta+\mathbf{u}_2$$
$$\left.\times(v_x/\beta c)(1-\cos^2\phi\sin^2\theta)^{\frac{1}{2}}]dz'\right|^2. \quad (59)$$

The integral is extended from $-l/2$ to $l/2$ (l is the total length of the undulator).

Introducing the following variables:

$$\xi=2\pi z'/l_0=\omega_0 z'/\beta c \quad (\omega_0=2\pi\beta c/l_0), \tag{60a}$$

$$\Omega_\theta=\frac{\omega_0}{1-\beta\cos\theta} \quad \left(\Omega_0=\frac{\omega_0}{1-\beta}\sim\frac{\omega_0 2\beta}{1-\beta^2}\sim\frac{2\omega_0}{1-\beta^2}\right), \tag{60b}$$

$$Z=\frac{\omega}{\Omega_\theta}\frac{e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}l_0\sin\theta\cos\phi}{2\pi\beta m_0c^2(1-\beta\cos\theta)}, \tag{60c}$$

the expression (59) takes the form

$$P_{k\omega}=C\left|\int_{-\pi l/2}^{\pi l/2}\left(\exp\left\{-iZ\sin\xi+i\frac{\omega}{\Omega_\theta}\xi\right\}\right)\times\left[\mathbf{u}_1\sin\theta+\mathbf{u}_2\sin X\frac{l_0e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}}{\beta^2m_0c^2 2\pi}\cos\xi\right]d\xi\right|^2, \tag{61}$$

where we have written $\sin X$ for $(1-\cos^2\phi\sin^2\theta)^{\frac{1}{2}}$ and C for $(e^2\beta^2/8\pi^3r^2c)(\omega/\omega_0)^2$. Under the integrand we expand

$$e^{-iZ\sin\xi}=\sum_{n=-\rho}^{\infty}J_n(Z)e^{-in\xi}, \tag{62}$$

and obtain the result

$$C\left|\mathbf{u}_1\sin\theta\sum_{n=-\infty}^{\infty}J_n(Z)\int\exp\left[i\left(\frac{\omega}{\Omega_0}-n\right)\xi\right]d\xi-\mathbf{u}_2\sin X\frac{l_0e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}}{\beta^2m_0c^2 2\pi}\sum J_n(Z)\times\int\exp\left[i\left(\frac{\omega}{\Omega_\theta}-n\right)\xi\right]\cos\xi d\xi\right|^2. \tag{63}$$

After an elementary trigonometric manipulation this becomes

$$C\left|\sum_{n=-\infty}^{\infty}J_n(Z)\left[\mathbf{u}_1\sin\theta\int\exp\left[i\left(\frac{\omega}{\Omega_\theta}-n\right)\xi\right]d\xi-\frac{\mathbf{u}_2}{2}\sin X\frac{e\mathcal{E}l_0(1-\beta^2)^{\frac{1}{2}}}{\beta^2mc^2 2\pi}\left(\int\exp\left\{i\left[\frac{\omega}{\Omega_\theta}-(n-1)\right]\xi\right\}d\xi+\int\exp\left\{i\left[\frac{\omega}{\Omega_\theta}-(n+1)\right]\xi\right\}d\xi\right)\right]\right|^2. \tag{64}$$

To discuss Eq. (64) we remark that the integrals

$$S_q=\int_{-l\pi/l_0}^{l\pi/l_0}\exp\left[i\left(\frac{\omega}{\Omega_\theta}-q\right)\xi\right]d\xi=\frac{2\sin[(\omega/\Omega_\theta)-q]l\pi/l_0}{(\omega/\Omega_\theta)-q} \tag{65}$$

$q=n,\quad n-1,\quad n+1$

are well known in diffraction theory. They show a large maximum in a narrow frequency range

$$|\delta\omega|<(l_0/2l)\Omega_\theta$$

and, of course, subsidiary maxima. It is seen that ω is restricted to small frequency bands centered at the "harmonics" of the "fundamental frequency" Ω_θ.

The energy radiated per unit area into the pth harmonic of Ω_θ is given by

$$\frac{1}{(1-\beta\cos\theta)^2}\left(\frac{\omega}{\Omega_\theta}\right)^2\frac{e^2\beta^2S_p^{\,2}}{4\pi^2r^2c}\left\{\sin^2\theta[J_p(Z)]^2+\tfrac{1}{4}\sin^2X[J_{p+1}(Z)+J_{p-1}(Z)]^2\cdot\frac{e^2\mathcal{E}^2l_0^2(1-\beta^2)}{4\pi^2\beta^2(m_0c^2)^2}\right\}. \tag{66}$$

The result obtained by putting $p=0$ deserves special mention. It is very small in our case. However, it is this term which would give the contribution of the Cerenkov effect to the total energy radiated if the undulator were filled with a medium of refractive index $\epsilon^{\frac{1}{2}}$. In this case, the velocity of light, c, where it represents the propagation of velocity of the radiation, has to be replaced by $c/\epsilon^{\frac{1}{2}}$ and the effect of this is that Z in Eq. (66) has to be replaced by $\epsilon^{\frac{1}{2}}Z$ and that the whole expression (66) is multiplied by $\epsilon^{\frac{1}{2}}$, while Ω_θ is now given by

$$\Omega_\theta=\omega_0/(1-\beta\epsilon^{\frac{1}{2}}\cos\theta). \tag{67}$$

The integrals S_0 give a contribution when $\omega/\Omega_\theta=0$, i.e., when

$$\beta(\epsilon)^{\frac{1}{2}}\cos\theta=1. \tag{68}$$

The angle θ as defined by Eq. (68) is real only when

$$\beta>1/\epsilon^{\frac{1}{2}}. \tag{69}$$

Thus, no contribution is obtained in free space. It is not intended to study the Cerenkov effect in this paper.

For the discussion of Eq. (66) first note that in case of the electric undulator we have always

$$e\mathcal{E}l_0\ll\pi mc^2. \tag{70}$$

We can replace $J_p(Z)$ by $Z^p/p!2^p$, and we note that the energy radiated into higher harmonics is small compared with that radiated into the fundamental frequency Ω_θ.

For the energy radiated into the fundamental ($p=1$) frequency band under angles ranging from 0 to θ we obtain

$$\frac{e^2\beta^2}{4\pi^2c}\int_0^\theta d\theta\int d\phi\left(\frac{\omega}{\Omega_\theta}\right)^2\left[\int S_p^{\,2}d\omega\right]\left\{\frac{\sin^4\theta\cos^2\phi}{(1-\beta\cos\theta)^4}+\frac{1-\sin^2\theta\cos^2\phi}{(1-\beta\cos\theta)^2}\right\}\frac{e^2\mathcal{E}^2(1-\beta^2)l_0^2\sin\theta}{4\cdot4\pi^2\beta^2(m_0c^2)^2}. \tag{71}$$

The integration over $d\omega$ may be carried out first. It

involves

$$\int S_p^2 d\omega = \int_{\omega/\Omega_\theta=1}^{\infty} \frac{4\sin^2[(\omega/\Omega_\theta)-1]l\pi/l_0}{[(\omega/\Omega_\theta)-1]^2} d\omega$$

$$= 4\frac{l\pi^2}{l_0}\Omega_\theta. \quad (72)$$

We are mainly interested in the result of the integration

$$(e^2/mc^2)\beta l\mathcal{E}^2(1-\beta^2)/8\pi.$$

$$\left\{\int_0^{2\pi} d\phi \int_0^{\theta} d\theta \left[\frac{\sin^5\theta\cos^2\phi}{(1-\beta\cos\theta)^4} + \frac{\sin\theta-\sin^3\cos\phi}{(1-\beta\cos\theta)^2}\right]\left(\frac{\omega}{\Omega_\theta}\right)^2\right\} \quad (73)$$

for small angles θ.

The frequency spread depends on the angle θ under which the radiation is received. Let us say we aim at a frequency spread of 10 percent. Then

$$d\Omega_\theta/\Omega_\theta = \beta\sin\theta d\theta/(1-\beta\cos\theta) \sim \beta\theta^2/(1-\beta) = 1/10,$$
$$\theta^2 \sim (1/10)(1-\beta^2)/2\beta. \quad (74)$$

We can put $\cos\theta=1$ in the integrand, and we find that the first integral can be neglected compared with the second. Integrating over ϕ we find that the energy radiated into a cone of opening θ is approximately 1/20 of the total energy as given by Eq. (27).

4. Correction for the Constant Velocity Orbit Approximation

It is now fairly easily seen how the calculation could be modified to include terms of order a^2l_0/v^2 arising from the constant velocity orbit approximation. The integrand of formula Eq. (73) is multiplied with

$$\exp\{i\omega[(a^2z'/2v^2)+(l_0a^2/4\pi v^2)\sin 2\omega_0 z'/v]\}.$$

For $\exp[i\omega(l_0a^2/4\pi v)\sin 2\omega_0 z'/v]$ we write

$$\sum_{r=\infty}^{\infty} J_r\left(\frac{l_0a^2}{4\pi v^2}\right)e^{2ir\xi}. \quad (75)$$

However, as the argument of the bessel functions is small, the series reduce to the terms $r=0$, which is unity, in our case. The term $\exp(i\omega a^2z'/2v^2)$ gives rise to a small change of the frequency Ω_θ which is now given by

$$\Omega_\theta' = \omega_0/[1+(a^2/2v^2)-\beta\cos\theta]. \quad (76)$$

IX. CONCLUSIONS

1. Production of Energy in the Infrared and Millimeter Band

The efficiency of energy conversion depends entirely on whether the energy is radiated coherently or not, and this, in turn, depends on whether the electrons can be so bunched that, seen from the L-system, they look no longer than half the wavelength of the radiation received in the L-system.

The undulator considered above delivers the equivalent of 3 quanta of maximum fundamental frequency per electron. In the millimeter band this corresponds to a total radiation of $2\cdot10^{-15}$ ergs/electron. Let the beam contain 1 coulomb/sec, then we have $\frac{3}{5}\cdot10^{19}$ electrons per second. The power radiated coherently by electrons which are suitably bunched depends on the number of electrons in the bunch and the nch repetition rate. Let us consider millimeter waves.

If we assume that is is possible to bunch all the electrons contained in one millimeter length of the beam into half that length, and to repeat a half-millimeter bunch every millimeter along the beam. The number of electrons/bunch is then

$$N = \tfrac{1}{5}\cdot10^8$$

and the power is $N^2\cdot2\cdot10^{-15}$ number of bunche/sec.

Power radiated coherently thus becomes

$$(\tfrac{1}{5}\cdot10^8)^2 2\cdot10^{-15}3\cdot10^{11} = 2.4\cdot10^{11} \text{ erg/sec} = 2.4\cdot10^4 \text{ watts},$$

while the power radiated incoherently would have been $1.2\cdot10^{-3}$ watts. However, if it were possible to bunch all the electrons contained in one-cm length of beam into $\frac{1}{2}$-millimeter bunches, we should get $2.4\cdot10^5$ watts radiated coherently.

This is the total energy radiated. The energy contained in a band of 10 percent frequency spread collected under an angle θ given by Eq. (74) is 1/20 of this, i.e., $0.6\cdot10^{-4}$ watt incoherent or $1.2\cdot10^4$ watts coherent radiation, on the second assumption and $1.2\cdot10^3$ watts on the first.

In the first case the power radiated coherently is $2\cdot10^7$ times the power radiated noncoherently from a uniform beam, while in the second case it is $2\cdot10^8$ times that quantity.

However, the total beam energy is only $1\cdot15\cdot10^6$ watts, so the energy radiated in the second case is 21 percent of the beam energy. We have obviously reached the limit of validity of our theoretical assumptions. The radiation reaction per electron is no longer negligible with respect to the external magnetic forces acting on the beam and should be considered in the dynamics of the electron.

We note that we may start with an undulator period $l_0=16$ mm, use a beam of 1.5 Mev and obtain $l_0'=4$ mm in the E-system. Then, for coherent radiation the bunches in the E-system must not be longer than 2 mm, while they should not exceed $\frac{1}{2}$ mm in the L-system. This seems just feasible.

The bunching system should be such that the speed variation within the bunch is small.

The maximum frequency emitted is given by

$$\nu = (l_0/v)1/(1-\beta).$$

Hence, the frequency spread due to velocity spread is given by

$$d\nu/\nu = -d\beta/(1-\beta).$$

It is seen that speed variations in a bunch lead to an amplified frequency spread of the emitted radiation.

2. Speed Monitoring for the Whole Beam

Speed monitoring is practicable if the emitted radiation can be detected and its frequency determined with sufficient accuracy. We are now thinking about noncoherent radiation, but it is easily seen that the energy involved is sufficient for the purpose. Even at millimeter wavelength the power is of the order of milliwatts, and much more if for higher beam energies. There are 3 quanta emitted per electron crossing an undulator of length of one meter with $l_0 = 1$ cm. A beam of one microamp, mean current contains $\frac{3}{5}\cdot 10^{13}$ electrons/sec, or $\frac{3}{5}\cdot 10^{7}$ per microsecond, and there will be about $2\cdot 10^{7}$ quanta emitted per microsecond, or 10^{6} in the 10 percent bandwidth considered above.

It seems convenient to use such an apparatus which can determine $1-\beta^2$ directly and a train of waves 100 long affords ample precision for wavelength measurement.

3. Speed Determination for Individual Particles

Here the figure of 3 quanta per electron seems rather small. It seems necessary to go to a greater length of undulator. An undulator 10 meters long would produce 30 quanta which could be detected by a scintillation counter.

X. ACKNOWLEDGMENTS

The author wishes to thank Professor F. Bloch and Professor L. I. Schiff for helpful suggestions and criticisms.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Electrical Breakdown over Insulators in High Vacuum

PAUL H. GLEICHAUF
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received October 2, 1950)

In the investigated range of 5×10^{-3} to 10^{-7} mm Hg, the breakdown voltage over insulators in high vacuum is independent of pressure. Currents of 10^{-11} to 10^{-8} ampere were observed in the region below breakdown voltage by detecting x-ray quanta with a Geiger-Mueller counter. Pinhole camera x-ray pictures revealed that practically all radiations originate from an area on the anode a distance from the insulator, with a weaker radiation coming from a ring immediately adjacent to the insulator. Current-voltage relationships as usually observed in these experiments indicate a roughness factor and an emitting area on the cathode similar to previous findings in vacuum gaps. Current bursts were observed which did not develop into complete breakdown. Oscilloscopic observations revealed that sometimes at breakdown over insulators the voltage on the test sample drops to 2.5 kv; in other cases it falls to less than 100 volts. The low voltage arc-like discharge extinguishes at a current of about one ampere for copper electrodes in contact with Pyrex glass. When a resistance in series with the test sample is increased to keep the maximum current below one ampere, no stable discharge is observed. As in a vacuum gap, the breakdown voltage over an insulator is increased by successive breakdowns. Part of this "conditioning" is permanent. The nonpermanent part is dependent on the state of the test sample prior to conditioning. The anode does not appear to influence conditioning. When the resistance in series with the test sample allows a discharge current above one ampere to flow, a fast conditioning usually occurs which results in a high permanent breakdown voltage.

INTRODUCTION

THE main limitation of high vacuum insulation is a spontaneous discharge, often called breakdown. The characteristics of high vacuum breakdown are high currents, preceded by very low pre-breakdown current; low voltage drops in the vacuum gap during breakdown; and in practically all cases the independence of breakdown voltage on pressure, provided that the pressure is below that at which a glow discharge may develop. Since the residual gas cannot supply a sufficient number of charged particles to carry the high breakdown current, gas or vapor must be released by the breakdown process.

If electrodes in a high vacuum are separated by an insulator, breakdown over the insulator, without its puncture, takes place at lower voltage than in a vacuum gap with the same separation of electrodes.

Preliminary work by Dr. M. J. Kofoid[1] showed that the breakdown voltage over a hollow glass cylinder which rests on a flat electrode is increased by more than 100 percent when one end of the insulator fits into a groove in the cathode. No similar effect is found for the opposite polarity. Breakdowns over a hollow glass cylinder fused to the cathode occur at higher voltages than when this glass piece rests on a well-polished electrode. Kofoid's tests with such metal-to-glass seals at the cathode indicate that the breakdown voltage changes little with the length of the insulator.

The writer made further investigations on this

[1] Unpublished work done at these Laboratories.

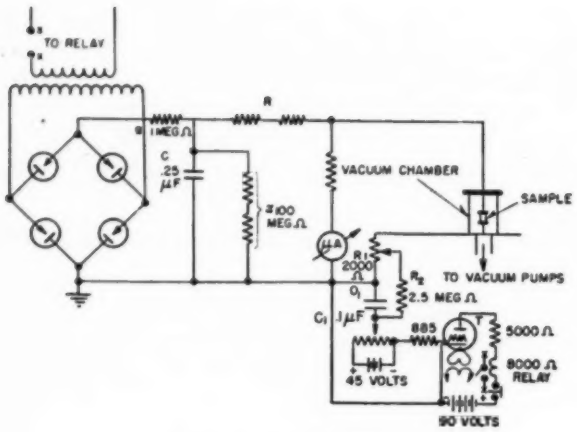


FIG. 1. Circuit diagram.

project after some work preliminary to the content of of this paper was done in cooperation with Mr. G. C. Thomas.

DESCRIPTION OF TEST EQUIPMENT

The majority of the tests were carried out in vacuum systems continuously evacuated by 4-inch oil diffusion pumps employing Litton oil. These systems had no cold traps. The glass cylinders of the vacuum chambers were about 18 inches high and 11 inches in diameter. Neoprene gaskets covered with Apiezon-N grease sealed the systems. The pressure was usually below 10^{-5} mm Hg, as indicated by an ionization gauge. For preparing sealed-off tubes, a vacuum system with mercury diffusion pumps and a liquid nitrogen trap was used. The tubes were baked at a temperature of about 400°C; during and after sealing the tube off, zirconium, which was not volatilized, served as getter. The pressures were about 10^{-7} mm Hg as indicated by an ionization gauge.

The source of the dc high voltage was a transformer with a full-wave bridge vacuum tube rectifier capable of delivering a maximum voltage of about 100 kv. The circuit diagram is shown in Fig. 1. At breakdown a thyratron (T) was fired, and this energized a relay which in turn opened the low voltage side of the high voltage transformer.

In most tests the grounded cathode was a copper disk 123 mm in diameter with rounded edge. The copper anode, which was 89 mm in diameter, was designed to prevent breakdowns in the vacuum space instead of over the insulator in cases where short lengths of insulators were used. Figure 2 shows its shape, which was arrived at by testing a model of the electrodes in a three dimensional electrolytic tank. In the sealed-off tube the electrodes were of 18–8 stainless steel. The electrodes were polished with emery paper, the finest of which was 4/0, then washed in water and acetone (cp). The fused quartz or 7740 Pyrex glass insulators were either hollow cylinders or rods whose flat surfaces were polished with carborundum powder 600 mesh. They were washed with soap and water, using a wetting agent (Aerosol), rinsed in water and acetone. The length of the insulators was usually ≈ 23 mm; the diameter of the rods was about 13 mm.

EFFECT OF VACUUM CONDITIONS ON BREAKDOWN

Should the breakdown voltage increase with decreasing pressure, one would suspect that ions of the residual gas are of importance in the breakdown mechanism. In a dynamic vacuum system, permitting measurements down to pressures of about 5×10^{-6} mm Hg, no effect of pressure on breakdown voltage was found up to a pressure of 5×10^{-3} mm Hg, at which pressure a glow discharge occurred.

Experiments were made with test samples in sealed-off tubes in which pressures of about 10^{-7} mm Hg were achieved. Later, these tubes were opend to the atmosphere and re-evacuated in the large vacuum system with oil diffusion pumps. The behavior of the test samples in both cases was substantially the same.

During breakdowns in the sealed-off tube it was observed that in successive tests less and less gas was developed, but no correlation was found between the breakdown voltage and the amount of gas released between any two successive tests. New breakdown tests were always made after released gases in the tube cleaned up, either because of readsorption or because of the pumping action of the ionization gauge. It is possible that part of the evolved gas came from the walls of the tube, because other tests showed that charged particles strike the walls of the vacuum chamber.

MEASUREMENT OF CURRENTS IN THE REGION BELOW BREAKDOWN VOLTAGE

Experimental Procedure

For the understanding of the breakdown mechanism it is of importance to know the origin of the charged particles which are essential for initiation of a breakdown. It is therefore natural to look for the existence of currents in the region below breakdown voltage and to study their magnitude and the mechanism of their origin. When this is done, it must be kept in mind that the magnitude of such currents does not necessarily have a direct bearing on the breakdown mechanism. However, should one find a critical current density close to breakdown, a direct relation between these currents and breakdown would be indicated.

Indirect measurements of prebreakdown currents were made by detecting x-ray quanta with a Geiger-Mueller counter which was placed outside the vacuum chamber. By this method only the current of electrons striking the anode was detected, and no leakage current interferes with the measurements. The Geiger-Mueller counter is sensitive even to short bursts of radiations, whereas most other devices do not respond to rapid current changes. Observed short time fluctuations in the number of detected counts were found to be within

the limits prescribed by statistical theory of detection of x-ray quanta. For the investigation of these statistical fluctuations, motion pictures were taken of the count indicators on the scaler. The pictures were taken at 8 frames per second and the time of exposure was 1/64 second.

Calibration of the Geiger-Mueller counter was made by using the same copper anode as in measurements with insulators, and a thermionically emitting tungsten filament cathode which gave currents readable on a microammeter. Such a calibration is justified because x-ray pinhole camera pictures revealed that practically all x-rays produced when an insulator bridges the gap have their origin at the anode. The efficiency of x-ray production[2] and absorption of x-rays by the glass[3] of the vacuum chamber may be calculated. In some vacuum gap tests with copper electrodes it was possible to read with a galvanometer currents at voltages close to breakdown, and equivalent x-rays were at the same time detected by the Geiger-Mueller counter. The accuracy of the calibration was within 15 percent. For tests in the sealed-off tube a correction of the calibration had to be made for an Fe target.

The voltage on the test sample was raised in steps of a few kilovolts until breakdown occurred. Counts were recorded at each voltage over a period of five seconds.

Results

Typical current *versus* voltage "curves" are shown in Fig. 2. Currents in the range from 10^{-11} to 10^{-8} were detected. The points are widely scattered, but there is a general trend of the data. Curves 1, 2, ···, 14 represent

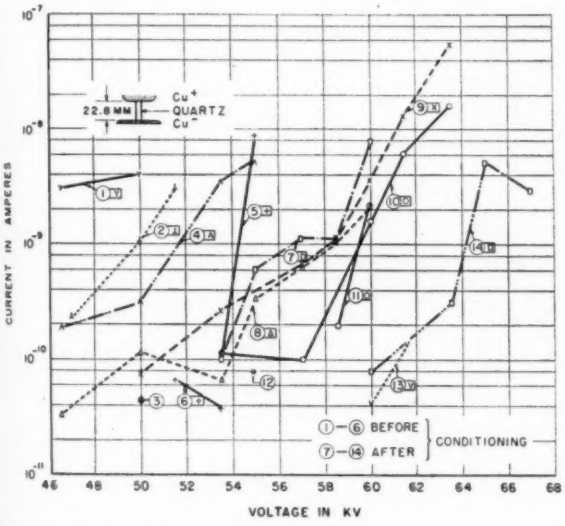


FIG. 2. Pre-breakdown current *versus* voltage for a test sample in a vacuum system with oil diffusion pumps. Each set of tests was terminated by a breakdown. The last point on each set represents the highest voltage for which breakdown did not occur. Curves 1, 2, ···14 represent successive sets of tests.

[2] R. T. Beatty, Proc. Roy. Soc. (London) **A89**, 314 (1913).
[3] E. W. Pike, J. Appl. Phys. **12**, 206 (1941).

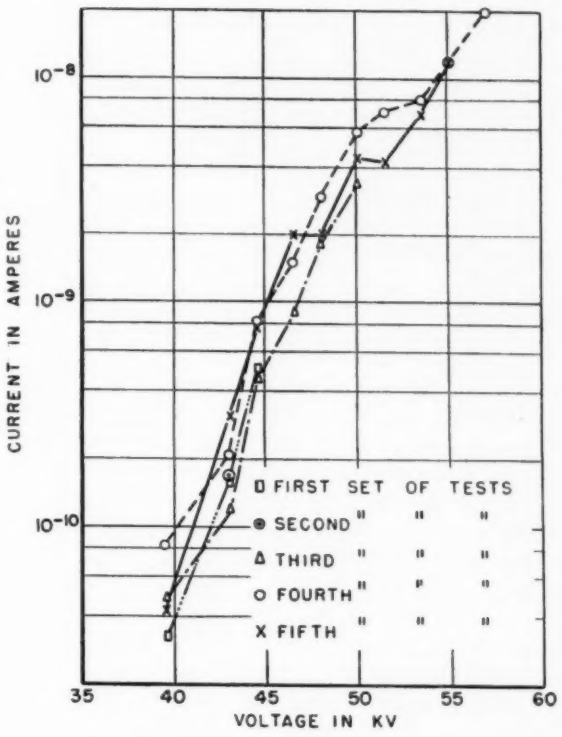


FIG. 3. Pre-breakdown current *versus* voltage. Test sample in sealed-off tube.

successive sets of tests. Curves 7 to 14 were taken after the test sample had been conditioned by successive breakdowns. A shift of the curves to higher voltages may be observed; that is, after conditioning, lower currents occurred at a given voltage. No critical pre-breakdown current was found. Sometimes breakdowns occurred after a current of about 5×10^{-11} amp and in other cases after a current of 5×10^{-7} amp was detected.

Because of the scattered points (Fig. 2), which were obtained by registering the number of counts over five-second time intervals, similar measurements were made with vacuum gaps using the same but freshly polished electrodes as used in previously mentioned tests. Similar scattering of points occurred. When the gap was conditioned by drawing current, but without breakdown, the scattering was appreciably reduced.

Further tests were made in the baked-out and sealed-off tube. The resulting curves of the conditioned test sample are shown in Fig. 3. This graph shows less scattering than occurs in Fig. 2.

Discussion

The reliability of measuring the electron current striking the anode from detection of x-ray quanta is supported by two facts; namely, (1) by statistical analyses of the data and (2) by the observed reduction of scattering of points on the current-voltage characteristic after conditioning (explained later) by drawing current or by baking. These observations are in agree-

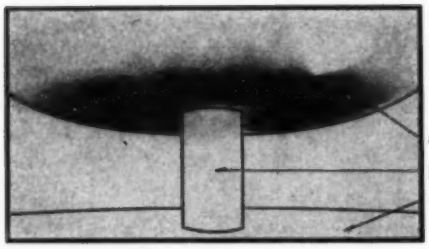


FIG. 4. X-ray pictures with pin-hole camera (negative). The black lines were added later to outline the test sample.

ment with the results of others who measured currents in vacuum gaps with a galvanometer.[4]

From the study of motion pictures of the scaler indicators when a constant voltage is applied to the test sample, one sees that there are three kinds of fluctuations in the number of detected counts. First one observes statistical fluctuations which are inherent in the detection process. The second kind of fluctuations, which occurs only occasionally, is that of a very high number of counts registered on a single frame of the film. This is attributed to sudden current bursts which do not develop into breakdown. The third kind, which probably accounts mainly for the scattering of the points in the current-voltage characteristics, and is observed with vacuum gaps as well as with insulators, is of sudden changes in current to a new "steady" value.

For comparison of the results with field emission, the emitting area on the cathode and the roughness factor, by which it is necessary to multiply the average field to obtain the maximum field at the cathode, were calculated from the equation,[5]

$$I=6.2\times10^{-6}\frac{\mu^{\frac{1}{2}}}{(\kappa+\mu)\kappa^{\frac{1}{2}}}\times F^{2}\times\exp\left(-\frac{6.8\times10^{7}\times\kappa^{\frac{1}{2}}}{F}\right),$$

where I is the current density in amperes per cm^2, $\mu\approx5$ ev, κ is the thermionic work function which was assumed to be about 4.5 ev, and F is the maximum field at the cathode. The data obtained with the test sample in the sealed-off tube were used.

The maximum field at the cathode is $F=(a\beta)V/d$, where V is the applied voltage to the test sample and d is the separation of electrodes (=5.5 mm). Here (a) is mainly determined by the presence of the insulator and (β) by roughness of the cathode surface. The value of $(a\beta)$ calculated from the curves in Fig. 3 is 435. If (a) is due only to the dielectric constant, which is around 4 for quartz, β is about 110. This value is in the range as that calculated from our data with vacuum gaps. From the work of Haefer[6] roughness factors of this magnitude may be estimated for a metal surface which was not heat treated. The calculated emitting area is $\approx1\times10^{-11}$ cm^2, which is of the same magnitude as areas obtained by others for high field emission.

[4] R. J. Piersol, Phys. Rev. **31**, 441 (1928).
[5] Stern, Gosling, and Fowler, Proc. Roy. Soc. (London) **A124**, 699 (1929).
[6] R. Haefer, Z. Physik **116**, 604 (1940).

PINHOLE CAMERA X-RAY PICTURES

It is not only of importance to know whether electrons strike the insulator and what field distribution exists, but for the previously mentioned method of current measurement, it is essential to know the energy of the electrons which produce the x-rays. Such investigations were made with pinhole cameras.

Experimental Procedure

For the investigation of the origin of x-rays and for an indication of their wavelengths two pinhole cameras with different filters were placed in the vacuum chamber at equal distance from the insulator. In one camera a film protected by 0.2 mm of paper was used which at 5 kv ($\approx$2.5A) absorbed about 99.4 percent of the radiations. The film in the other camera was protected by two aluminum foils each about 5×10^{-5} cm thick with a maximum absorption of 75 percent at the characteristic wavelength of aluminum at about 7A and an absorption of about 5 percent for 5-kv x-rays.

Results

A typical picture (negative) taken on a film protected with paper is shown in Fig. 4. The test sample is outlined by heavy black lines which were drawn on the negative. During the exposure of this film many breakdowns occurred over fused quartz at voltages between 50 and 85 kv. Such pictures reveal that practically all x-rays have their origin on the anode around the insulator. Less blackening occurs in a ring immediately adjacent to the insulator.

On the films which were protected by thin aluminum foils there are faint indications of radiations coming from regions corresponding to the insulator. These radiations are more intense closer to the cathode.

Discussion

The pattern produced on the x-ray films suggests that electrons are deflected by the insulator which has acquired a negative charge. The slight blackening in a ring adjacent to the insulator could be due to soft x-rays which are produced during breakdown when only a low voltage remains across the test sample. This is suggested by comparison of pictures obtained with films protected by paper and by thin aluminum foils. Since the electrodes were designed after a model which proved to provide a uniform field, electrons could not have been deflected by a peculiarity of the field. Nor is the deflection due to a positive charging of the glass cylinder of the vacuum chamber. This was confirmed by an experiment in which a large grounded cylindrical shield was placed between the glass wall of the vacuum chamber and the test sample; in this case the same kind of x-ray picture was obtained as before.

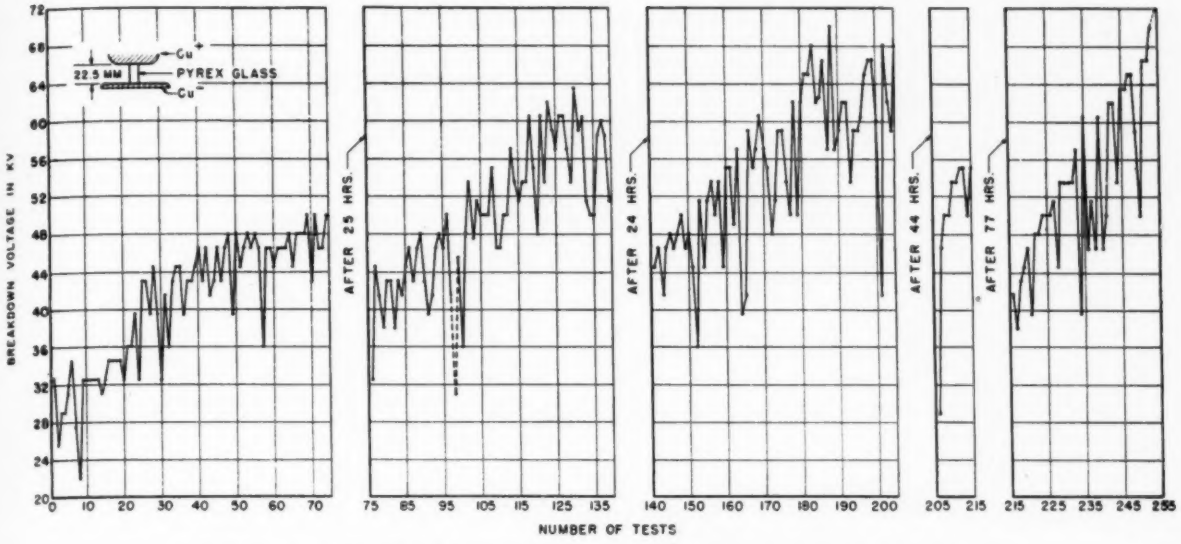


FIG. 5. Fluctuations of breakdown voltage during conditioning by successive breakdowns. One-minute time intervals between starts of tests.

The faint indication of radiations coming from regions corresponding to the insulator could be due to soft x-rays of an average wavelength ≈ 2.5A.

OSCILLOSCOPIC OBSERVATIONS OF BREAKDOWN

Experimental Procedure

For oscilloscopic observations of current at breakdown, the vertical deflection plates of the cathode-ray tube were connected directly to a shunt in the discharge circuit. Continuous sweep was used. The insulators were glass rings 22.5 mm high. Tests of the circuit using sphere gaps in air showed that the observed trace gives a true indication of the current.

The voltage across the test sample at the time of breakdown were calculated from the observed currents in tests where the electrodes were separated by a 2-mm high glass ring. Short insulators were used to make the ratio of the voltage drop to breakdown voltage as large as possible because it was assumed that such a voltage drop is practically independent of breakdown voltage. Each picture of the trace on the oscilloscope representing current at breakdown over the insulator was compared with two or three pictures of current at breakdown of sphere gaps in air. The spacing of the spheres was adjusted so as to achieve breakdowns in air at the same voltage at which breakdown in vacuum over the insulator had occurred. The tests were made at voltages above the permanent value obtained after conditioning.

Results

Sometimes at breakdown over insulators the voltage across the test sample falls to 2.5 kv; in other cases the voltage drops to less than 100 volts. The probable error in the value to which the voltage falls is about 160 volts. Breakdowns occur in these tests between 10 and 15 kv.

The low voltage arc-like discharge extinguishes at a current of about one ampere (between 0.6 and 1.7 amp) for copper electrodes in contact with hollow glass cylinders. This is true irrespective of the value of the breakdown voltage and of the value of the series limiting resistance. With the experimental setup used for these measurements no observation of the prebreakdown current in vacuum was possible.

Copper electrodes in a vacuum gap of few millimeters separation were tested for extinguishing currents. The results are similar (0.7 to 1.5 amp) to those observed over insulators.

Discussion

When the series limiting resistance was increased to keep the maximum current below one ampere, no stable discharge was observed. Apparently, the discharge takes place with a current of one ampere but lasts for a very short time (in the range of 10^{-5} to 5×10^{-5} second with the given arrangement). When a condenser of about 300 $\mu\mu$f is connected in parallel with the test sample, the discharge lasts longer, without a change of breakdown voltage. This suggests that the discharge currents are due to the discharging of the capacity of the test sample and of stray capacities of the circuit.

CONDITIONING

Experiments with vacuum gaps reveal that a higher breakdown voltage is achieved by successive breakdowns. A similar "conditioning" is observed in breakdowns over insulators.

Several kinds of studies on the conditioning phenomenon were made, namely, the general behavior, the part played by the electrodes, the part played by the

insulator, and the effect on conditioning of the circumstances under which the arc-like discharge extinguishes.

Experimental Procedure

In all described tests the voltage of the power supply was slowly raised at a uniform rate (about 5 kv/sec) up to breakdown voltage, at which time the relay interrupted the low voltage side of the power supply. The limiting resistor R in series with the test sample was, except where specified, ≈ 10 megohms. Glass and fused quartz insulators were used.

Long series of successive tests were made to investigate the fluctuations in breakdown voltage and their general trend. Other tests were made to investigate the part played by the electrodes in conditioning. This was done by comparing the average breakdown voltage of a conditioned test sample before the contact area between the insulator and one or both electrodes was changed with a similar average value afterwards. It was of further interest to find out what role the insulator plays in conditioning. First a vacuum gap with copper electrodes was conditioned by successive breakdowns. Then the electrodes were separated and, without opening the vacuum system, a fresh insulator, longer than the original separation of the electrodes in the vacuum gap, was inserted. The electrode separation in the vacuum gap was such that breakdowns occurred at substantially higher voltages than over the insulator. During conditioning of the vacuum gap, the insulator was kept within the vacuum system in a brass box wrapped in a lead foil. The values for the average breakdown voltage over the insulator before and after conditioning were compared. Current observations on an oscilloscope revealed that with a series limiting resistor (R in Fig. 1) larger than a critical value (R_{crit}) no stable discharge occurs. This is due, as explained before, to the extinguishing of the discharge at a current of about one ampere. Studies were made to determine the effect of a resistor $R<R_{crit}$ on conditioning. Measurements were made with copper, 18–8 stainless steel, and aluminum electrodes.

Results

(1) During successive breakdowns it was observed that, as the test progresses, a general trend toward higher breakdown voltages as well as a diminishing rate of rise is indicated. This is shown in Fig. 5, which shows also the large fluctuations in breakdown voltage. The phenomena are best illustrated by the first group of tests (Numbers 1 to 75). One sees further that the breakdown voltage rises more rapidly in each successive series of tests, indicating that faster conditioning takes place than in the preceding series. One concludes therefore, that conditioning depends on the previous treatment of the test sample. Between the groups of tests, no voltage was applied for the time intervals specified. When tests were resumed, breakdown voltages were considerably lower than at the end of the previous sequence of tests. This decay of breakdown voltage during the time in which no voltage was applied was dependent on the preceding treatment of the test sample. As the time intervals between successive series of tests were lengthened, the average breakdown voltage, upon resumption of testing, reaches a lower limit (V_0) which is usually higher than the voltage at which breakdowns occur before conditioning. V_0 is practically identical with the average breakdown voltage of a conditioned test sample which has been exposed to atmospheric pressure. The length of exposure to atmospheric pressure is immaterial. From these tests it was concluded that for a new test sample V_0 is already reached 24 hours after conditioning. Conditioning occurs also with the high voltage connected to the test sample continuously. Sparks occur less and less frequently at a constantly applied voltage. When the voltage is raised, breakdowns are again frequent; but the probability that a breakdown will occur decreases with time of application of voltage. After reaching a sufficiently high voltage, the breakdowns become very frequent and no further conditioning can be observed.

(2) If the insulator is moved on the cathode or on both electrodes, the breakdown voltage is on the average (average of several test runs) 15 percent lower than before moving; but these average values measured after the moving of the insulator on the cathode, or on both electrodes, are still higher than the average of the first five breakdown voltages measured initially on the test sample after it is assembled. It is further observed that, after the relative position of the insulator and the cathode is changed, a faster conditioning takes place than with a newly assembled test sample. Decay measurements were made to determine whether or not the lower breakdown voltage, in the case where measurements are made 60 to 120 seconds after conditioning, is due merely to a decay of breakdown voltage with time. These measurements show that, at breakdown voltages observed and for the short time intervals used, the ordinary decay of breakdown voltage with time does not account for the reduction noticed after moving the insulator. A strong adherence of the insulator to the electrodes is observed after the conditioning of the test sample. The increase in breakdown voltage which results from conditioning was shown not to be due to the adherence of the insulator to the electrodes. The role which the anode plays in conditioning was also investigated. Although the results of these tests are less conclusive than the tests with the cathode, no effect of the anode on conditioning is found.

(3) The average voltage of the first five breakdowns over a new insulator inserted between the conditioned electrodes for several test runs is about 25 percent lower than the breakdown voltage V_0 measured after a permanent conditioning is achieved. In all tests, except one, the average breakdown voltage was increased by

conditioning. The case where the breakdown voltage was not increased after the inserted insulator was conditioned might be similar to the rarely observed cases where V_0 is practically identical with the initial average breakdown voltage.

In some tests the conditioned glass insulator was removed without breaking the vacuum, and the electrodes were brought again to the original separation of the vacuum gap. Lower breakdown voltages than after the vacuum gap was assembled were now observed. The following test suggests that the decrease is not due to a lowering of the work function by an electro-positive layer of sodium on the cathode. Rings of fused quartz, low in alkaline metal content, were used as insulators. The results were similar to those with glass. The lowering of breakdown voltage in the vacuum gap might be attributed to small insulating particles which remain on the electrodes after the insulator is moved.

(4) With a resistance below the critical value, a fast conditioning is usually observed which results in a high permanent breakdown voltage V_0. This breakdown voltage remains high when a resistance above the critical value is connected in series with the test sample, but is lowered when the test sample is exposed to atmospheric pressure. With 18–8 stainless steel electrodes and $R<R_{crit}$, the fluctuations in breakdown voltage are sometimes large and then little conditioning is observed.

Discussion

It is difficult to explain the part which the cathode plays in the permanent conditioning during breakdowns over insulators so long as no complete theory has been developed for conditioning in vacuum gaps. When currents in the region below breakdown voltage are measured, a fatigue is observed. Millikan and Eyring[7] tried to explain this phenomenon in vacuum gaps by the assumption that sharp microscopic peaks on the cathode, from which electrons are emitted, are battered down or rounded off by ion bombardment. It is not yet clear if such a theory can partly explain conditioning by successive breakdowns, because it is usually[8,9] observed that after conditioning of a vacuum gap by successive breakdowns, the currents prior to breakdown were increased; but a simultaneous occurrence of several effects resulting in conditioning by successive breakdowns may account for the latter phenomenon.

As far as fatigue of currents prior to breakdown is concerned, one should mention a paper by Haefer.[5] He observed with an electron microscope that for experiments in a residual gas atmosphere of argon at a pressure of about 10^{-5} mm Hg the sharpness of the emitting points may be increased as prebreakdown current flows. He was able to melt these points by increased field emission currents. It seems possible, therefore, that these points sometimes become so fine that the local current density reaches a value high enough to melt them.

Millikan and Eyring suggested further that impurities lowering the work function of the cathode are removed. Similarly, it seems that in our case impurities may also be removed from the insulator. A further possibility might be that in some cases insulating particles and insulating oxide patches present on the cathode are removed. Positive charges altering the field at the cathode could accumulate on such insulating layers.

One could expect that particles coming from the anode take part in the initiation of breakdown as suggested by experiments of Chambers.[10] Ahearn,[9] however, did not find such an effect. Similar experiments in which it would be possible to eliminate the effect of the glass walls of the tube on conditioning of a vacuum gap could be of interest.

The nonpermanent conditioning may be due partly to a removal of adsorbed gases from the insulator; the time decay could be due to readsorption of the residual gas. Charging effects may also play a part. Neither charging effects of the insulator nor removal of adsorbed gases seem to explain the faster conditioning and higher breakdown voltage experienced with a resistance smaller than the critical value which allows a current above one ampere to flow at breakdown.

ACKNOWLEDGMENT

The writer is especially grateful for the advice and many helpful suggestions given by the late Dr. R. C. Mason during the course of this work.

[7] R. A. Millikan and C. F. Eyring, Phys. Rev. **27**, 51 (1926).
[8] W. H. Bennett, Phys. Rev. **37**, 582 (1931).
[9] A. J. Ahearn, Phys. Rev. **50**, 238 (1936).
[10] C. C. Chambers, J. Franklin Inst. **218**, 463 (1934).

An Electrical Method for Investigating the Nature and Behavior of Small, Airborne, Charged Particles

J. H. DANIEL AND F. S. BRACKETT
Laboratory of Physical Biology, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland

(Received July 19, 1950)

A theoretical and experimental analysis of the current-voltage (dc) characteristic of a parallel-plate condenser through which the small, airborne, charged particles of an aerosol are passed continuously shows that a charge-weighted distribution of radius (or radius squared)-charge ratio may be obtained when certain experimental conditions are satisfied. Examination of such a distribution for an aerosol (particles less than a micron in diameter) before and after passage through a small (20-liter) settling chamber shows that the classes of particles in the distribution that are easiest to collect in the condenser are also the classes that disappear most rapidly in the chamber. It is pointed out that the methods employed would become less tedious and more powerful when used with a dispersion of constant-size particles, and that they are well adapted for certain fundamental investigations of aerosol properties.

1. INTRODUCTION

IN order to evaluate properly the results of experiments on the physiological effects of inhaled aerosols,[1] it is necessary to have information on such physical characteristics as the mass and charge distributions, and the state and rate of aggregation of the particles composing the aerosol. For particles below 0.5 micron in diameter, whose presence may well be crucial to the experiment,[2] measurement of such characteristics is particularly difficult.

Simplification can sometimes be obtained by the use of particles of a narrow range of sizes.[3] When this is possible, the uncertainties of mass distribution effects are eliminated, and the three independent distributions of charge, mass, and charge-mass ratio are reduced to one of charge alone. For this simplified case, the method to be described here allows an absolute determination of the charge distribution, if aggregation can be neglected.

In the most general situation of independent and changing charge, mass, and charge-mass ratio distributions in the presence of aggregation, our method determines a charge-weighted distribution of radius (or radius squared)-charge ratio for a given state of aggregation. Such distributions, taken for different states of aggregation and at different concentrations, can give considerable information about the nature and behavior of the aerosol particles.

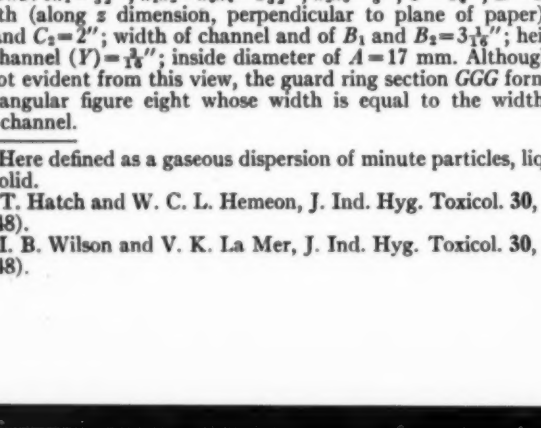


FIG. 1. Apparatus for measuring currents caused by aerosol particle charges. Positively/negatively charged particles are collected by rectangular condenser plates C_1 and C_2 when potentials E_1 and E_2 are positive/negative. Condenser dimensions are as follows: $ox_1=1\frac{7}{32}''$; $x_1x_2=x_3x_4=1\frac{23}{32}''$; $x_2x_3=\frac{7}{8}''$; $T=4\frac{3}{4}''$; $D=1\frac{5}{8}''$; width (along z dimension, perpendicular to plane of paper) of C_1 and $C_2=2''$; width of channel and of B_1 and $B_2=3\frac{1}{16}''$; height of channel $(Y)=\frac{3}{16}''$; inside diameter of $A=17$ mm. Although it is not evident from this view, the guard ring section *GGG* forms a rectangular figure eight whose width is equal to the width of the channel.

The method involves, essentially, an analysis of the current-voltage (dc) characteristic of a parallel-plate condenser through which the aerosol is passed continuously. Theoretical investigation of particle trajectories in the condenser (Sec. 4), prompted by the desire to obtain some characteristic of the aerosol involving particle size and/or charge, showed that a curve can be derived from the original current-voltage curve (as in Sec. 5, Figs. 6–8) which, for certain easily satisfied conditions, has ordinates proportional to $\bar{e}dN/d(r/e)$, where N is the number of particles having a radius-charge ratio r/e, and $\bar{e}$ is the average charge of these particles. The abscissas will be proportional to r/e. This derived curve is thus a charge-weighted distribution of r/e (for particles small compared with the mean free path of air molecules, r/e must be replaced by r^2/e). Since the proportionality constants can be determined, size distributions for the charged particles can be calculated from assumed relationships between r and e. The quantity $\bar{e}dN/d(r/e)$ is quite sensitive to the effects of aggregation, diffusion to walls, and gravity settling in reducing the charge and number of the particles and in increasing their effective size.

The purpose of this article is to describe this method and to demonstrate its validity. It is pointed out that certain defects in our apparatus could be wholly or partially eliminated, and that our procedures could be improved in several respects. Although we have not attempted this, the results we have obtained appear to be sufficiently quantitative to justify further analysis. Most of this is done in a separate article.[4]

[1] Here defined as a gaseous dispersion of minute particles, liquid or solid.

[2] T. Hatch and W. C. L. Hemeon, J. Ind. Hyg. Toxicol. **30**, 172 (1948).

[3] I. B. Wilson and V. K. La Mer, J. Ind. Hyg. Toxicol. **30**, 265 (1948).

[4] J. H. Daniel and F. S. Brackett, manuscript submitted for publication to A.M.A. Archives of Industrial Hygiene and Occupational Medicine.

TABLE I. Current-voltage data, taken with the apparatus of Fig. 1, for a silica aerosol.

Potential of 1st and 2nd plates (E_1 and E_2)	Current to 2nd plate (I_2)	Repeat run (I_2)
	Sample passed through 20-liter settling chamber	
45 volts	7.6×10^{-14} amps (10:45 A.M.)	7.9 (11:05)
90	16.0	15.8
180	21.7	21.9
360	18.6	17.7
675	11.2	11.0
− 45	− 4.5 (10:55)	− 4.5 (11:15)
− 90	− 9.2	− 9.6
−180	−14.4	−15.2
−360	−13.3	−13.9
−675	− 8.5	− 9.4
	Sample direct from exposure chamber	
22.5	6.8 (11:30)	7.0 (11:50)
45	17.1	16.4
90	31.0	30.1
180	36.3	35.0
360	25.5	24.2
675	14.7	15.2
− 22.5	− 3.8 (11:40)	− 3.8 (12:00)
− 45	− 9.1	− 9.0
− 90	−19.2	−19.3
−180	−26.0	−25.7
−360	−20.5	−20.3
−675	−12.0	−13.0 (12:10)

The figures in parenthesis are times of the observations. Silica concentration at the end of the experiment = 52 mg/m³. Flow through 2.6 m³ exposure chamber = 280 l/min. Sample flow through condenser = 1.45 l/min. Temperature constant at 78°F. Relative humidity constant at 47 percent.

2. APPARATUS

Sensitivity and accuracy being the primary requirements of the current-measuring system, with portability highly desirable, the FP-54-electrometer tube in a DuBridge-Brown compensated circuit[5] was used. In Fig. 1, the aerosol passes from a cylindrical delivery tube A, through a fish-tail taper section T, into a duct D of rectangular cross section, and enters the double,[6] parallel-plate condenser at 0, which is taken as the origin of a rectangular coordinate system. An electrostatic field F_1 is applied between condenser plates B_1 and C_1, and F_2 between B_2 and C_2, by maintaining plates B_1 and B_2 at appropriate battery potentials E_1 and E_2, respectively. Collector plates C_1 and C_2, each surrounded by a grounded guard ring system G, are maintained at ground potential by being connected individually, through gold-tipped polystyrene-insulated switches, either to ground directly, or to the FP-54 grid and an appropriate resistor R. Using the FP-54 as a potential indicator, the current in R due to collection of charged aerosol particles can be calculated from the nulling voltage V required between R and ground. Any one of a series of values of R may be switched in to cover a range of current magnitudes, or a polystyrene-insulated brass condenser (approximately 10 micro-microfarads) may be used in place of R to measure extremely small currents by a rate-of-charge method. All switches and resistors in the measuring system are enclosed in a metal box containing the FP-54 and a drying agent. Complete electrostatic shielding is provided for appropriate parts of the apparatus.

A plane, midway between the condenser plates, divides the apparatus into a top half and a bottom half, each consisting of a Formica plate properly milled to form the aerosol channel. The brass collectors and guard-ring system are set in a polystyrene plate milled to receive them. The polystyrene plate does not contact the Formica; the guard ring serves as a bridge between them. This construction allows easy disassembly for cleaning.

As a source of aerosol, generators and animal exposure chambers, which have been described previously in the literature,[7] were used. Because one or two hours may be required to complete a set of measurements, it is essential that the characteristics of the aerosol remain constant over such a period. To this end, and to allow more nearly reproducible control, several modifications in apparatus and procedure were incorporated. These are discussed in a separate article.[4]

The sample of aerosol for analysis by the apparatus of Fig. 1 was continuously withdrawn from the exposure chamber through a 17-millimeter (i.d.) glass tube, the

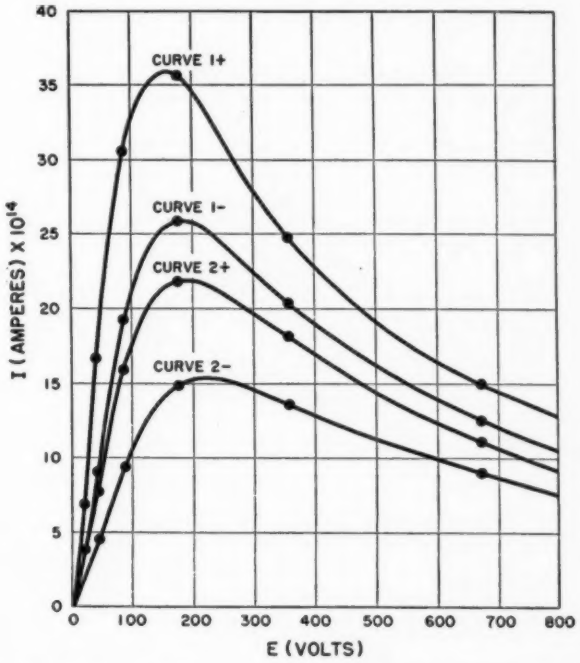


FIG. 2. Plots of the current-voltage data of Table I, showing the excess of positive over negative charge and the decrease in charge after passage through a settling chamber. Curves 1 are for silica samples direct from the exposure chamber; curves 2, for samples which, in addition, have passed through the settling chamber. The + and − signs denote currents of positive and negative charges, respectively. A detailed analysis of these curves is undertaken in Sec. 6a.

[5] L. A. DuBridge and H. Brown, Rev. Sci. Instr. 4, 532 (1933).
[6] While not necessary in principle, the double condenser allows experimental checks on operating conditions (see Sec. 5).
[7] Dautrebande, Alford, Irwin, Mitchell, Thompson, Weaver, and Wood, Arch. intern. pharmacodynamie 80, 388 (1949).

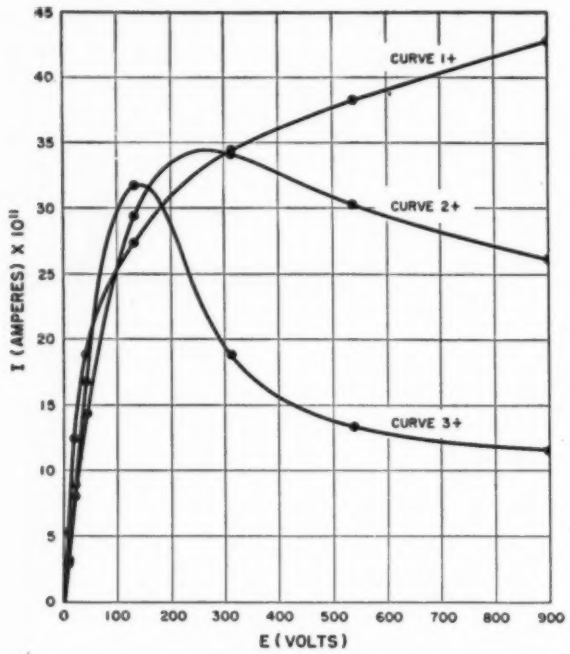


FIG. 3. Current-voltage characteristics for a distilled water aerosol at three positions along the generator delivery tube. Curve 1 was taken at a position 55 cm from the generator jets, curve 2 at 165 cm, and curve 3 at 340 cm. Evaporation to smaller, easier-to-collect particles is suggested by the decrease of charge associated with the hard-to-collect particles being collected at high voltages (on the right), while the lack of a compensating increase in charge associated with the small, easy-to-collect particles collected at low voltages (on the left) suggests aggregation or loss to the tube walls. Generator temperature was 14.6°C; room temperature 27.0°C. Evaporation of water from aerosol particles was practically complete at a point 220 cm from generator jets. Further analysis of these curves is presented in Sec. 6b.

condenser of Fig. 1, a flow meter, and a 40-liter air reservoir, by means of a vacuum pump normally used with the Mine Safety Appliances electrostatic precipitator. This pump was run by a constant, low speed motor, and the aerosol flow (usually of the order of 1 l/min) regulated by varying either the constriction in the rubber tube to the air reservoir, or the opening of a parallel outlet tube to air.

3. REPRESENTATIVE DATA

A representative sample of current-voltage data, taken under the conditions described in the previous section for a silica aerosol, is shown in Table I. The current measured at the second plate is recorded for various voltages applied to both plates, both before and after passing through a 20-liter bottle. Two runs were made in order to check the steadiness of conditions, negligibility of polarization effects, etc.

This data is plotted in Fig. 2. It is seen that the current to the second plate rises as increasing voltage allows more and more charged particles to be collected. It reaches a maximum when the rate of gain of charge produced by the collection of more of the relatively "hard-to-collect" particles equals the rate of loss of charge produced by "easy-to-collect" particles falling to the first plate. Finally, it decreases at high voltages, which bring most of the charge to the first plate.

Such an interpretation has been verified by any number of combinations of measurements of currents and voltages on both plates, and it appears, from these, that the efficiency of collection approaches 100 percent. Whether a large or small fraction of the particles stick to the plate on collection is debatable. Certainly a large number do, for after several hours operation, the currents become erratic at high voltages, and the condenser plates must be cleaned of a thin but highly visible coating of particles.

In addition to the shapes of the curves of Fig. 2, and the relative magnitudes of currents due to positive and to negative charges, an interesting feature is that the currents of both positive and negative charges are substantially reduced after passing through the 20-liter settling jar. Similar characteristic changes are evidenced by the curves of Fig. 3. These are current-voltage characteristics taken in a similar manner at various positions along a glass delivery tube leading directly (with no intervening exposure chamber) from an aerosol generator charged with distilled water. From this sequence it appears that, as the particles move along the tube, they become much easier to collect, as might be expected if their size is reduced by evaporation. However, there appears to have been a large number of easy-to-collect particles close to the generator; and as we go along the tube, the total charge associated with these particles does not increase to compensate for the decrease of total charge associated with the hard-to-collect particles. This suggests neutralization of charge by aggregation, or loss to the tube walls.

To place such ideas on a more quantitative basis it becomes necessary to develop a theory covering the motion of these charged particles in the space between the condenser plates.

4. THEORY OF METHOD

The taper of Section T (Fig. 1), the length of duct D, and the velocity v of aerosol flow may be made such that parallel, streamline flow exists in the region $0<x<x_4$. The transition from streamline to turbulent flow for a rectangular section[8] occurs at a Reynolds number $4Sv\delta/p\mu$, where S = cross-sectional area, p = perimeter, v = average fluid velocity, δ = fluid density, and μ = fluid viscosity. Thus, substituting the values of the constants for air, the permissible values of aerosol velocity for streamline, parallel flow are given by $v<80(p/S)$ in the cgs system of units, provided T tapers slowly enough, or D is long enough. For a cross section $3\frac{1}{16}\times\frac{3}{16}$ inches, values found experimentally practical, we have $v<360$ cm sec^{-1}.

For such parallel, streamline flow, the velocity distribution in the xy plane (see Fig. 1) is parabolic, as given by $v_x=(4v_m/Y^2)(Yy-y^2)$. Thus, v_x, the velocity in

[8] W. H. McAdams, *Heat Transmission* (McGraw-Hill Book Company, Inc., New York, 1942), second edition, p. 124.

the direction of flow (x), is a function of the position between condenser plates (y) only. Midway between the plates, v_x has a maximum value $v_m=\frac{3}{2}v$, where v is the average velocity; at either plate ($y=0$, $y=Y$), it has the value 0. It is assumed that the z dimension of the collector is small compared with the z dimension of the rectangular section, so that, over the collector, the value of v_x is practically independent of this dimension.

Because we are concerned with particles well under 45 microns in radius,[9] the terminal velocities of the particles produced by the force of gravity are insufficient to give turbulent motion.[10] Calculation for the most extreme conditions of our experiments shows that the electrostatic forces produced on the particles by the field are several orders of magnitude less than the limiting value for departure from streamline motion. Thus, in discussing particle motion, the only departures from Stokes' equation which need be considered are those resulting when the particle radius r becomes so small as to be comparable to the mean free path L of the air molecules. These departures may be taken into account by using Millikan's semi-empirical equation for resistance to particle motion, which includes and connects the two theoretically justified expressions for regions of small and large L/r.[11] The y component of the equation of motion of a spherical aerosol particle of charge e, mass m, and radius r, in the presence of an electrostatic field along the y direction then becomes

$$\frac{dv_y}{dt}+av_y-b=0; \quad a=\frac{6\pi\mu r}{m[1+A'(L/r)]},$$

$$b=\left(\frac{Fe}{m}\pm g\right), \quad A'=(A+Be^{-C(r/L)}), \qquad (1)$$

where g is the acceleration of gravity, and A, B, and C are constants with values 1.23, 0.41, and 0.88, respectively.

The term $\pm g$ in the expression for b of Eq. (1) may be eliminated by orienting the apparatus so that vectors g and v_x are parallel, in which case g gives rise to an entirely negligible term in the expression for v_x. If this is not done, the significance of g may be determined experimentally by observing the difference in data taken with the force of gravity parallel ($+g$), and then antiparallel ($-g$) to v_y.[12]

The dispersing effect of brownian motion "random walk," neglected in Eq. (1), may be calculated from the expression[11] $\bar{y}^2=(2kT/c\mu)t$, where $\bar{y}^2$ is the mean-squared displacement in the y direction, during time t, of a particle in a fluid at temperature T. In this expression, k is Boltzmann's constant, and $1/c\mu$ is the steady velocity of drift under unit driving force,[13] which, again using Millikan's empirical formula, equals $[1+A'(L/r)]/[6\pi\mu r]$. Letting t be the *average* time of transit through the condenser (about 3 seconds in most experiments), calculation shows that $(\bar{y}^2)^{\frac{1}{2}}$ is roughly 0.3 mm for $r=0.01$ micron, and 0.3 cm for $r=0.001$ micron or 10 angstroms. Since the x velocity distribution along the y direction is parabolic, transit times for particles close to the condenser plates are larger than 3 seconds, and it becomes evident that the regions close to the condenser plates will be robbed of particles of 0.01 micron radius and below.

Calculations similar to those of Wilson,[14] using average particle charges and charge densities as determined in Sec. 6, indicate that if all particles were of like sign, there might also be some loss of the smallest particles to the walls because of the mutual electrical repulsion. This loss is probably inappreciable in our experiments, where comparable numbers of particles of both signs are present. Rough calculations of the field produced by the particles when positive and negative particles are separated by an applied field to opposite condenser plates shows this perturbation to the applied field to be negligible under the conditions of our experiments. Calculation of electrostatic image forces shows them to be significant only at extremely small distances from the condenser plate surfaces and, therefore, negligible for our purposes.

Experimentally, no phenomena which could be interpreted as being due to ionization or electrical breakdown were observed. Unfortunately, a direct experimental test of the justification of neglecting brownian motion, electrostatic repulsion, and image forces at the walls, based on the observation of negligible current to collectors C in the absence of field F, is insensitive in the presence of nearly equal positive and negative charges.

The solution of Eq. (1), for $v_y=0$ at $x=0$, when $t=0$, is

$$v_y=(b/a)(1-e^{-at}). \qquad (2)$$

For small particles, experimental conditions are usually such that the exponential term in t in Eq. (2), which arises from the "inertial" term dv_y/dt in (1), may be neglected. Physically, this means that a particle reaching the collector plate will have traveled practically its entire excursion in the y direction at a terminal velocity in that direction determined by Millikan's semi-empirical equation, the time required for acceleration to that velocity being negligible. Experimentally,

[9] Some idea of the range of particle sizes involved in the experiments may be obtained from Fig. 1 and reference 4 of Dautrebande *et al.*, Occupational Med. **5**, 508 (1948). The particles probably range from almost molecular dimensions to less than a micron. See also Fig. 12 of text.

[10] J. M. Dalla Valle, *Micromeritics, the Technology of Fine Particles* (Pitman Publishing Corporation, New York, 1948), second edition, p. 21.

[11] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938).

[12] For $r=0.1$ micron, $e=$unit electronic charge, $F=$field produced by 20 volts (least value used in most experiments) between plates separated $\frac{3}{16}$ inch, and particle density $d=2$, $g=0.1Fe/m$, and is nearly negligible in the expression for b, as long as F and g act for the same period of time. See, however, reference 18.

[13] $kT/c\mu$ is the diffusion coefficient of the particles in air.

[14] I. B. Wilson, J. Colloid Sci. **2**, 271 (1947).

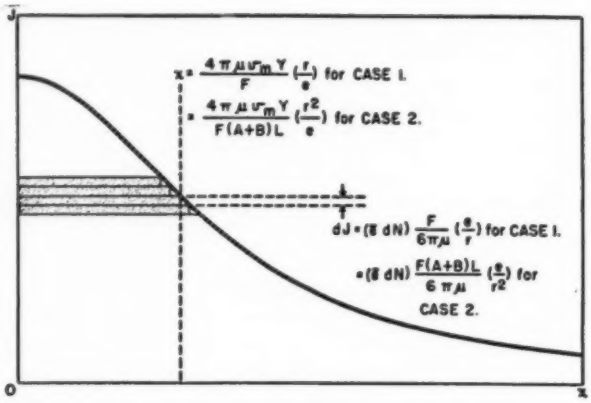


FIG. 4. A plot of total current density J to the collector plates C and guard rings G of Fig. 1 as a function of distance from the origin 0, for an arbitrary applied field F. This curve results from the summation of the constant individual r/e class contributions dJ, having the magnitudes shown for Case 1 or for Case 2. For Case 1, $L/r \ll 1$, and for Case 2, $L/r \gg 1$, where L is the mean free path of air molecules and r the aerosol particle radius.

the unimportance of this inertial term can be verified by noting that the current to the second collector at zero potential remains negligible for any value of potential applied to the first plate. Analytically, it can be shown by solving Eq. (2) for the distance s traveled in the y direction, and showing that, for particles reaching collector C, this distance is inappreciably affected by omission of the exponential term. Thus, for $s=0$ when $t=0$,

$$s=(b/a)t-(b/a^2)(1-e^{-at}). \tag{3}$$

The second term on the right-hand side of Eq. (3) is the inertial term. Plotted against t, it is found to have initially (at $t=0$) the same slope (b/a) as the first term. With increasing t, however, the slope of the first term remains constant, while that of the second (inertial) term approaches zero as a limit, the term itself approaching a maximum value b/a^2. For the inertial term to be negligible, then, $(b/a)t \gg (b/a^2)$ for $t \gtrsim x/v_m$. This criterion becomes

$$ax/v_m \gg 1 \quad \text{or} \quad r^2 \ll 9\mu x/2v_m d(1+A'L/r), \tag{4}$$

where d = density of particle material.

Calling the term in parenthesis in Eq. (4) unity, for the moment, using 360 cm/sec as the safe upper limit of v_m to avoid turbulence in the condenser cited above, taking x equal to the guard ring width of 1.35 cm as the least favorable distance involved, and $d=3$ gcm^{-3} as not likely to be exceeded, we find $r^2 \ll 100$ microns. Actually, it was necessary to make v_m nearly two orders of magnitude less than the critical value for turbulence, in order to collect the particles with reasonable values of field. This further increases the limiting value of r. For such magnitudes of r, the term in parenthesis in Eq. (4) is indeed very close to unity; for very small values of r, it variesas $1/r$. Hence, if Eq. (4) is satisfied for any value of r, it is satisfied for all smaller values of r. Thus, for our experiment, where few particles were greater than a micron in diameter, we are justified in replacing Eq. (2) by

$$v_y=b/a=Fe(1+A'L/r)/6\pi\mu r, \tag{5}$$

when the apparatus is oriented so that F is perpendicular to g.

The time for a particle to travel the distance y thus becomes $t_y=y/v_y$, while the time to travel a distance x is

$$t_x=\frac{x}{y^{-1}\int_0^y v_x dy}=\frac{3Y^2x}{v_m(6Yy-4y^2)}.$$

Equating t_x and t_y, we may solve for the distance x traveled by a particle starting at $t=0$, from the position $x=0$, $y=y$, and arriving at the point $x=x$, $y=0$:

$$x=\frac{4\pi\mu v_m(3Yy^2-2y^3)}{FY^2[1+A'(L/r)]}\frac{r}{e}.$$

The maximum value of x, occurring for particles starting from $y=Y$, is

$$x_m=[4\pi\mu v_m Y/F(1+A'L/r)](r/e). \tag{6}$$

We have assumed that $F=0$ for $x<0$, and $F=$a constant for $x>0$. These conditions cannot be strictly realized. A complete analysis of the consequences of the resulting "edge" effect on the relations we shall derive would be difficult. However, its first-order effect would be an effective increase in the value of x by an amount proportional to the channel depth Y. Because Y is small in comparison with x_2, this discrepancy is probably unimportant, where measurements with the second plate C_2 are concerned, as long as B_1 and B_2 are at the same potential.

Let us take first the case for $L/r \ll 1$, designate it Case 1, and focus our attention on a given class of particles having a value of r/e lying between r/e and $(r/e)+d(r/e)$. The current density to the collector (i.e., at $y=0$) up to the point $x=x_m$, resulting from the particles in this class, is

$$dJ=-v_y d\rho=-v_y \bar{e} dN=-(F/6\pi\mu)(e/r)\bar{e}dN, \tag{7}$$

where $d\rho$ is the charge density, $\bar{e}$ the average charge per particle, and dN the number per cm^3 of particles with r/e between r/e and $r/e+d(r/e)$. The negative sign expresses the fact that an increase in current density corresponds to a decrease in aerosol charge density. Equation (7) is valid in this simple form only because v_y is assumed constant [Eq. (5)] for this class of particle, and dN, which is assumed uniform throughout the aerosol originally, remains unchanged in the channel (wherever these particles are found) as long as the drift velocity v_y is constant and the gas flow laminar and parallel.

Equation (7) shows that the current density, dJ, due to an incremental r/e class, is constant over the collector

plate up to $x=x_m=(4\pi\mu v_m Y/F)(r/e)$. Figure 4 shows how the summation of such incremental contributions for all r/e classes present gives rise to a plot of total current density J *vs* distance x along the condenser.

Making use of Eq. (6), simplified for Case 1 to the expression given without subscript in the preceding paragraph, Eq. (7) can be written

$$\frac{dJ}{d(\ln x)}=x\frac{dJ}{dx}=\frac{r}{e}\frac{dJ}{d(r/e)}=-\frac{F}{6\pi\mu}\bar{e}\frac{dN}{d(r/e)}=-k_1\bar{e}\frac{dN}{d(r/e)}. \quad (8)$$

Thus, if a curve like that of Fig. 4 is plotted on a logarithmic scale in x, its slope will be proportional to the number of particles of a given r/e class, weighted by the average charge per particle of that class.

When applied to Case 2, $L/r\gg 1$, the same type of argument that led to Eq. (8) for Case 1 now gives

$$\frac{dJ}{d(\ln x)}=-k_2\bar{e}\frac{dN}{d(r^2/e)};\quad k_2=\frac{F(A+B)L}{6\pi\mu}, \quad (9)$$

where N is now the number per unit volume, and $\bar{e}$ the average particle charge of particles classified according to their values of r^2/e. We now have a charge-weighted distribution of r^2/e rather than one of r/e.

The interpretation of the $[dJ/d(\ln x)]$ *vs* x curve as a charge-weighted distribution of r/e or r^2/e (both being αx) in Eqs. (8) and (9) requires that the area under this curve, as well as that under the J *vs* x curve of Fig. 4, should be proportional to the volume charge density ρ of the aerosol particles. This may be shown for Case 1 as follows:

$$\int_0^\infty -\frac{dJ}{d(\ln x)}dx=\int_0^{J_{\max}} xdJ$$

$$=\frac{4\pi\mu v_m Y}{F}k_1\int_0^\infty \bar{e}\frac{dN}{d(r/e)}d\left(\frac{r}{e}\right)$$

$$=\tfrac{2}{3}v_m Y\int_{r/e=0}^{r/e=\infty}\bar{e}dN.$$

The second and fourth expressions above are unchanged for Case 2; they represent the total current collected per unit plate width z, and are therefore equal to $\frac{2}{3}v_m Y\rho$.

When absolute magnitudes of r and e are involved in interpretations from experimentally determined curves of $dJ/d(\ln x)$ *vs* x, the transition region $L/r\doteq 1$ cannot easily be treated. However, many conclusions, involving only comparisons between aerosols, can be drawn regardless of the range in particle size, provided the limiting conditions considered above are satisfied.

Thus, plots of $dJ/d(\ln x)$ *vs* x, for a sample, taken at various times, will be sensitive to aggregation effects. If it is assumed that all particles, or a certain fraction of them, adhere upon collision, they will increase in effective r and decrease in e[15] as time goes on. This results in (1) shifting the curve to the right toward the large x of "hard-to-collect" particles of large r/e (or r^2/e), and (2) decreasing the area under the curve. Any information obtained solely from the area magnitudes could be obtained as well from simpler measurements of total current carried by the aerosol flow. However, additional information can be obtained by consideration of the shape of such curves. These considerations are developed further in reference 4.

It is not unreasonable to suppose that soon after dispersal the charge on the particles might be proportional to some power of radius,[16] i.e., $e=\bar{e}=\alpha r^\beta$, α and β being constants. For such a case, $r/e=r^{1-\beta}/\alpha$, and Eqs. (8) and (9) become, respectively,

$$\frac{dJ}{d(\ln x)}=-k_1\bar{e}\frac{dN}{dr}\frac{dr}{d(r/e)}=-k_1\frac{a^2r^{2\beta}}{1-\beta}\frac{dN}{dr} \quad (10)$$

for Case 1, and

$$\frac{dJ}{d(\ln x)}=-k_2\frac{\alpha^2r^{2\beta-1}}{2-\beta}\frac{dN}{dr} \quad (11)$$

for Case 2.

For $\beta=1$ in Case 1, and $\beta=2$ in Case 2, there is only one r/e and one r^2/e class, respectively, so that departures of the curve of Fig. 4 from a rectilinear plot are occasioned only by (1) the departure of L/r from either of the two limitations $1\gg L/r\gg 1$, (2) the average or statistical nature of the charge-radius relation. The reversal in sign at these two critical values of β signifies that, for greater values of β, particles with larger r are collected at smaller x; for β-values smaller than the critical values, particles with larger r are collected at larger x. The $dJ/d(\ln x)$ *vs* x curve becomes a size distribution curve for $\beta=0$ (charge constant) in Case 1; in Case 2 it becomes a distribution of r^2.

If size distributions are known, the applicable one of Eqs. (10) or (11) may be used to calculate curves of $dJ/d(\ln x)$ *vs* x for various values of β, provided the value of e for any single value of r is known (i.e., provided α is known). Alternatively, size distributions may be calculated from experimental curves of $dJ/d(\ln x)$ *vs* x, assuming a reasonable value of e for any value of r. Comparisons between such derived and experimental curves are an aid in determining the applicability of assumptions of the type $e=\alpha r^\beta$ (see Sec. 6a).

Such assumptions have a bearing on the effects of settling out due to gravity and diffusion. Thus, if the

[15] This will be true, on the average, if comparable quantities of + and − charge are present. Since the area under the curve is proportional to charge density, aggregation will fail to decrease it only if all particle charges are of the same sign—a case rarely encountered. In this extreme case, the effect of aggregation would be a change in shape and shift along the x axis.

[16] W. B. Kunkel, Phys. Rev. **78**, 91 (1950). In general, any such relation can hold only statistically, or for a certain fraction of particles, and the effective values of both α and β will vary in the course of time if aggregation takes place.

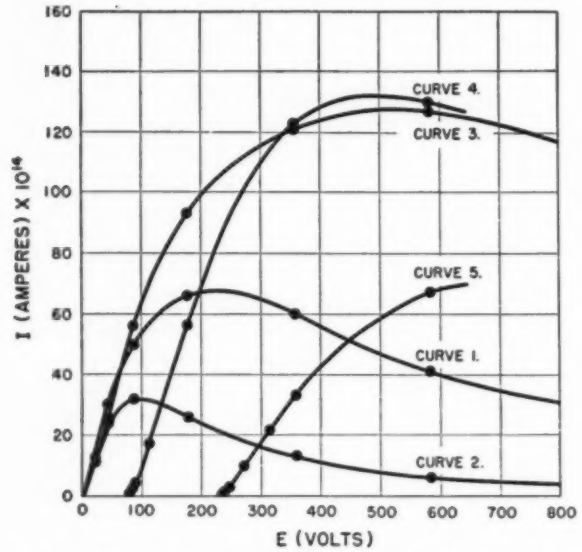


FIG. 5. Current-voltage characteristics for a silica aerosol. SiO_2 concentration at end of run = 46 mg/m³. Sample flow = 1.70 l/min (average velocity = 7.64 cm/sec). Curve 1: Current to first plate as a function of voltage on first and second plates (first plate used as a first plate). Curve 2: Current to second plate as a function of voltage on first and second plates (second plate used as a second plate). Curve 3: Current to second plate as a function of voltage on second plate; first plate held at ground potential (second plate as a first plate). Curve 4: Current to second plate as a function of voltage on second plate; first plate held at −45 volts. Curve 5: Current to second plate as a function of voltage on second plate; first plate held at −135 volts. The failure of curves 4 and 5 to intercept the axis at +45 and +135 volts, respectively, can best be accounted for by a constriction in the flow pattern near the entrance to the region between the first condenser plates (see text).

average size of particles collected at small x were found to be larger than that of those collected at large x, then $\beta >$ the critical values, the large particles removed by gravity are particles of small r/e (Case 1) or small r^2/e (Case 2), and the $dJ/d(\ln x)$ *vs* x curve is shifted to the right, or in the same direction as the shift due to aggregation. On the other hand, smaller particles found at small x ($\beta <$ critical values) would indicate that settling caused by gravity would cause the curve to shift to the left, opposing the shift due to aggregation. Since diffusion removes the smaller particles more rapidly, it tends to oppose gravity in shifting the curve to the right or left.

5. RESULTS ILLUSTRATING FUNCTIONING OF APPARATUS

With the apparatus illustrated in Fig. 1 it is, of course, impossible to obtain a complete curve of J *vs* x as in Fig. 4 for a single voltage or field. However, it follows from the expressions given in connection with the curve of Fig. 4 that a curve like that of Fig. 4 may be constructed from data obtained with constant x, but varying F, or varying v_m, by proper scale transformation. For varying F, this amounts to plotting Fx as abscissa, and J/F as ordinate; for varying v_m, x/v_m as abscissa and J as ordinate.

Such a plot for the second collector plate of our apparatus, using a silica aerosol, is shown in Fig. 6; the data from which it was plotted is shown as curve 2 in Fig. 5. Since the plate currents of curve 2, when reduced to a per unit area basis, are averages over a considerable range in x, the plate x dimension is drawn to scale in Fig. 6 for each voltage (or field F) used in plotting Fig. 6, and the curve drawn in so that the average current per unit area over the extent of the plate gives the observed average value.

Curve 1 of Fig. 5 gives the current-voltage characteristic found for the first plate of the apparatus; and curve 3 gives that for the second plate used as a first plate, i.e., with the first plate maintained at ground potential. Current readings for a given voltage were taken for each of the three curves before proceeding to the next voltage. The values plotted in Fig. 5 are averages of two successive runs whose consistency was comparable to that exhibited in Table I.

Curve 2 of Fig. 7 is the curve of Fig. 6 with the abscissa plotted on a logarithmic scale. Curves 1 and 3 of Fig. 7 are similar plots of the data of curves 1 and 3 of Fig. 5. The slopes of curves 1, 2, and 3 of Fig. 7 are plotted as curves 1, 2, and 3 in Fig. 8.

Figure 9 shows curves of the type of those of Fig. 7 for the first plate used as a first plate and the second plate used as a second plate for three different aerosol velocities, the Fx axis values being reduced by the ratio of the velocities so that, according to our theory, the curves should coincide.

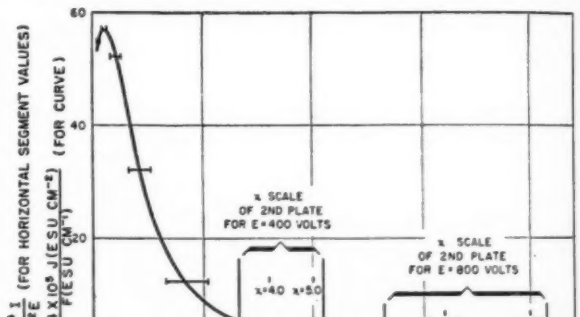


FIG. 6. Experimental data of curve 2 (second plate used as second plate) of Fig. 5 plotted so as to give the form of the theoretical curve of Fig. 4. I = current in amperes to the collector plate area of $3\frac{13}{16}$ in.²; E = potential difference in volts between condenser plates separated by $\frac{3}{16}$ in. Plots constructed in this manner allow the variation of current per unit area along the length of the condenser to be determined for any constant field, although the data was taken by varying the field and measuring the current at only a single position (that of the collector plate) along the condenser length. This is the first step necessary to convert the original current-voltage data to a charge-weighted distribution of (radius)^n-charge ratio. The negative slope, or drop of the left-hand portion of the curve toward the origin, could not occur if the limiting conditions assumed in obtaining Fig. 4 were satisfied. It can be explained as a result of the combined effects of diffusion, gravity, and a flow pattern over-concentrated at the center of the channel in creating a relative deficiency of particles in the channel region adjacent to the condenser plates.

All these curves are for positively charged[17] silica particles, with the condenser oriented so that gravity opposed the collection of particles. Curve 1 of Fig. 10 was taken in this manner, but in curve 2, gravity aided particle collection.[18] It would be possible to make a similar test of the effects of diffusion, using varying lengths of tubing, with flow direction vertical to eliminate gravity effects.

The following features of Figs. 5–10 require interpretation:

1. The rapid falling-off toward the origin of curves of the type of Fig. 6 (in contrast, the theoretical curve of Fig. 4 intercepts the y axis at a finite maximum value, with asymptotic direction parallel to the x axis). This results in the extreme left portion of curves like those of Fig. 8 falling below the axis.
2. Lack of agreement in Figs. 7 and 9 between curves made with first plate as first plate and with second plate as second plate.
3. Better agreement in Figs. 7 and 9 between curves made with second plate as first plate and second plate as second plate, with discrepancies mainly in the easy-to-collect and hard-to-collect regions of low and high voltage, respectively.
4. Tendency of higher velocity curves in Fig. 9 to lie above and to the right, especially in the easy-to-collect region.

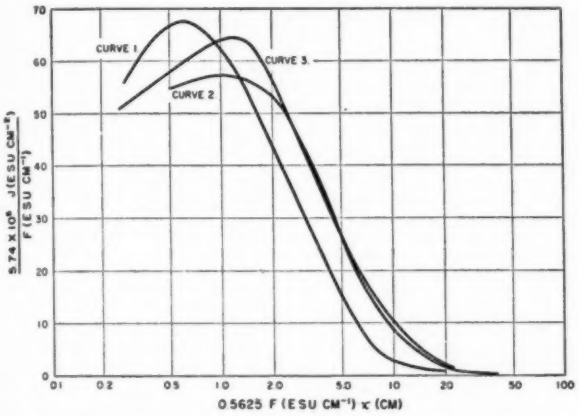


FIG. 7. Curves obtained by plotting the type of curve shown in Fig. 6 with the abscissas on a logarithmic scale. Curve 2 (data for second plate used as a second plate) is the curve of Fig. 6 with its abscissas on a logarithmic scale. This is the second step in converting the original current-voltage data of Fig. 5 into a charge-weighted distribution of (radius)^n-charge ratio. Curves 1 (first plate as a first plate) and 3 (second plate as a first plate) are similar curves for the data of curves 1 and 3 of Fig. 5. Had all the conditions imposed in Section 4 (such as absence of gravity, diffusion, and edge effects, and of departures from parallel, stream-line flow) been satisfied, these three curves should have coincided.

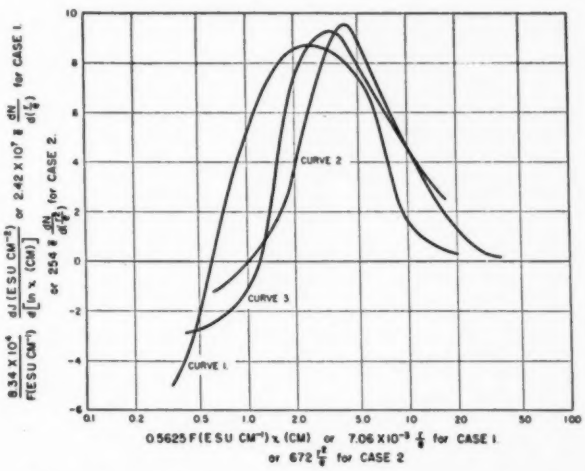


FIG. 8. Charge-weighted distribution of (radius)^n-charge ratio, on logarithmic scale. Curves 1 (first plate used as a first plate), 2 (second as second), and 3 (second as first) result from the similarly numbered curves in Figs. 5–7, and are obtained from the corresponding curves of Fig. 7 by plotting the slopes of these curves (their abscissas being expressed in log units in computing the slopes). This is the third and final step in obtaining the desired charge-weighted distribution. As in Figs. 6 (had curves 2 and 3 been shown there) and 7, the test of how well the limiting conditions imposed in Section 4 are satisfied lies in how nearly the three curves coincide. The negative ordinates at the left, which are, of course, impossible in an actual distribution, result from the negative slope discussed in connection with Fig. 6.

If a current-voltage characteristic of Fig. 5 approached the origin as a straight line, or as a curve with only downward curvature, the falling-off toward the origin in Fig. 6 could not occur. Thus, this falling-off (feature 1 above) reflects abnormally low current densities for weak fields, or abnormally high current densities for somewhat stronger fields. Since the greater part of the current at weak fields is contributed by particles already close to the channel wall on the collector side, a deficiency of particles in this space, or a surplus of particles in the channel mid-section, is indicated. Such a deficiency is caused, as we have seen, by gravity (Fig. 10, reference 18); diffusion to the walls (Section 4) undoubtedly contributes also. In addition, it is possible that departures from the postulated parallel lines of flow may aggravate this situation.

The relatively good agreement between curves 2 and 3 of Fig. 7, where the same plate, subject to the same flow conditions, was used as a first plate, and then as a second plate, and the relatively poor agreement between these curves and curve 1, where a different plate was used, may be taken as an indication that the flow conditions for the two plates are different. In fact, by the assumption of a *vena contracta* near the condenser entrance, resulting in a diverging flow in the dimension of the channel over the first plate, and a less diverging flow over the second plate, the differences among the curves of Fig. 7 and of Fig. 8 can be accounted for qualitatively.[19]

[17] Data for negatively charged particles were similar; comparisons are made in Sec. 6. Also see reference 4.

[18] The criterion given in reference 12 for negligible gravity effect is not strictly applicable in this experiment, because gravity separation in the aerosol occurred in a 100-cm distance (including tube lead of 17-mm inside diameter) before the region of the electrostatic field F was reached. The slower moving aerosol (v_m=9.7 cm sec) near the top and bottom walls will be robbed of larger particles, some of which carry (as will be shown) high charge. The region near the bottom wall will be replenished, to a certain extent, by large particles from the center region. The entire effect can be reduced by shortening the tube or spiralling the aerosol through it, or it can be eliminated by orienting the apparatus with flow in a vertical direction. For all other curves presented here, gravity separation was effective over about 50 cm.

[19] Some of the difference between curves 2 and 3 is undoubtedly due to the "edge" effect discussed in Section 4; however, this

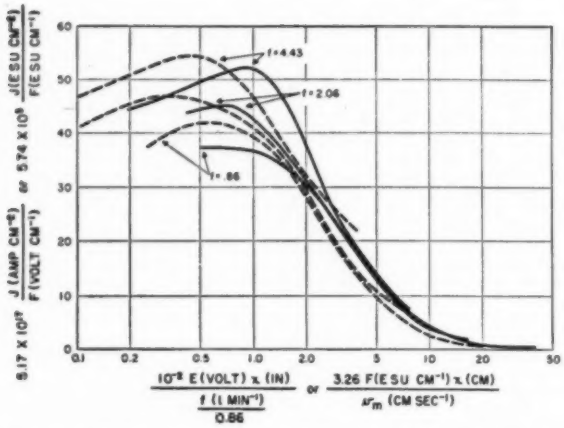


FIG. 9. Curves showing current density per unit field as a function of Fx/v_m for various values of F (electrostatic field), x (distance along condenser), and v_m (maximum velocity of aerosol through condenser). Broken line curves are for first plate used as a first plate; solid curves are for second plate used as a second plate. Maximum velocity v_m (cm sec^{-1}) $=\frac{3}{2}v$ (average velocity) $=6.74f$, where f is the aerosol flow through the condenser in liters per minute. For a given velocity, these curves are analogous to those of Fig. 7. When such curves are taken for various velocities, the method of plotting used here should cause them to coincide, provided again that the limiting conditions imposed in Section 4 are satisfied. The discrepancy between curves for the second plate and those for the first plate, here and in Fig. 7, and the relatively good agreement between curves for the same (second) plate, whether used as a first or as a second plate in Fig. 7, make it appear that flow conditions about the two plates are different, in agreement with the inference from curves 4 and 5 of Fig. 5. The slight displacement of the present curves to the right with increasing velocity could then be a result of a change in this flow pattern with velocity.

The current-voltage characteristics of the second plate, taken with the first at -45 and -135 volts, are shown as curves 4 and 5 of Fig. 5. With parallel flow, these curves should intercept the x axis at approximately $+45$ and $+135$ volts, respectively. The fact that the intercepts are found at about 1.6 times these values can best be explained as the consequence of a diverging flow pattern. For average flows of 0.86 and 4.43 liters per minute, this factor was 1.7 and 1.5, respectively, as might be expected if higher velocity displaced the *vena contracta* further down the channel.

From these observations, it would appear that in Fig. 9 the discrepancies between first and second plates are in large measure due to nonparallel flow conditions, while the discrepancies between curves for the same plate produced by different velocities (especially at curve extremities) are due chiefly to varying particle losses through gravity and diffusion, and in smaller measure to flow conditions. It seems reasonable to conclude that all the features of our experimental curves can be explained in terms of our theory, if the perturbing effects of gravity, diffusion, nonparallel flow, and edge effect are included. It is apparent that the first of these effects could be eliminated, and the remainder greatly reduced by proper experimental procedure and design.

The remainder of this paper is devoted to determining how much further we can carry the discussion of the curves of Section 3 (Figs. 2 and 3) with the aid of the analysis given in Sections 4 and 5.

6. APPLICATION OF METHOD TO SILICA AND WATER AEROSOLS

(a) Charge-Mass Relationship for Silica Aerosols

The considerations of Section 5 indicate that more nearly parallel flow conditions exist at the second plate than at the first. Data taken with the second plate used as a second plate are still affected by inclusion in the field of the nonparallel region around the first plate. But, because it has the advantages of greater resolution in the x dimension, less sensitivity to edge effect, greater ease of construction of curves like that of Fig. 6, and better coverage of significant portions of the curves at large Fx before reaching excessive voltages, this method was adopted for further investigations.

The charge-weighted distribution curves resulting from the data presented in Table I and Fig. 2 are shown in Fig. 11, plotted on a logarithmic scale in x. For both positive and negative particles, the decrease in height and shift of maximum, after passage through the 20-liter bottle, are in accord with our discussion of the

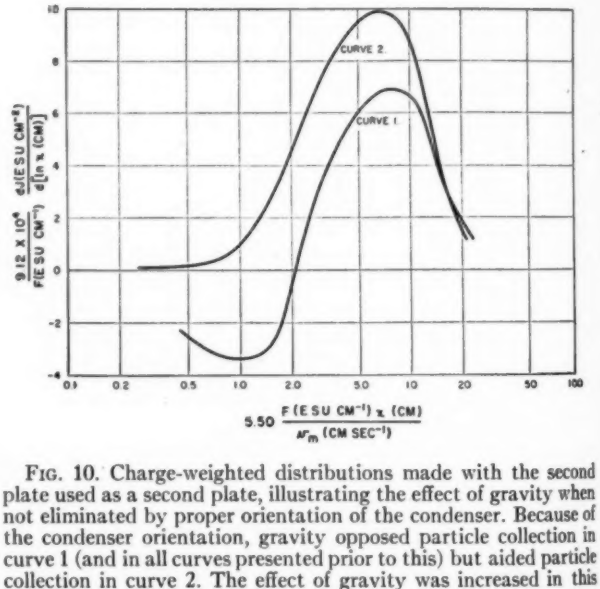


FIG. 10. Charge-weighted distributions made with the second plate used as a second plate, illustrating the effect of gravity when not eliminated by proper orientation of the condenser. Because of the condenser orientation, gravity opposed particle collection in curve 1 (and in all curves presented prior to this) but aided particle collection in curve 2. The effect of gravity was increased in this experiment by doubling the length of the inlet tube to the condenser (from 50 to 100 cm). For a fixed condenser orientation, the effects of diffusion might be demonstrated in a similar fashion by comparisons between curves taken for two different lengths of tubes leading to the condenser. The reduced time available for these two effects (gravity and diffusion) to exert their influence is probably responsible for the increase in current with increasing velocity so noticeable in the easy-to-collect (low voltage) range of the curves of Fig. 9. The aerosols for these curves, and for all curves presented hereafter, were produced with the modified apparatus described in Sec. 2, and with no animals in the exposure chamber. All experiments prior to this one were done with the unmodified apparatus and with animals in the chamber.

would seem to be small. The effect would be somewhat reduced by extending the existing guard-ring system into the region of negative x, and placing a grounded guard strip opposite this extension and in front of B_1.

effects of aggregation, gravity, and diffusion given in Sec. 4. Such considerations are amplified in reference 4.

It was characteristic of all curves for silica that the curve maximum for negative particles was found to the right of that for positive particles, although the mean abscissas of the curves for positive and negative particles (plotted *vs* x rather than $\ln x$) were about the same. Such a situation could result from association of the most probable value of r/e (or r^2/e) with smaller particles for the positive particles. This would be in agreement with a reported tendency of positive charges to be associated with smaller particles under certain conditions.[20] No such effect was observed with sodium chloride aerosols, where curves for positive and negative particles were closely symmetrical.[4]

No significant differences were found in such charge-weighted distributions taken when the silica was dispersed within an hour of its grinding and those taken days or weeks later.

To investigate the applicability of the relation $e=\alpha r^\beta$, discussed in Section 4, for the sample of Fig. 11, the curves of Fig. 12 were constructed. Published work[16] on the charges of larger particles of this type indicates that for these, charge varies roughly as the square of the radius. It is reasonable to assume that an appreciable number of the smallest of the much smaller particles considered here carry only unit electronic charge. For $\beta=0$ and $\beta=1$ for Case 2, the smaller particles fall to the left in the easy-to-collect region; hence, our assumption of unit electronic charge at the easy-to-collect limit is probably a good one.

It is evident that the value $\beta=0$ gives a range of sizes which is improbably narrow, and a curve of improbable shape for large r. Furthermore, calculation of the aerosol mass concentration M from the $\beta=0$ curve gives a value of 0.9 mg/m³ compared with 52 mg/m³ as measured with the electrostatic precipitator. By the assumption of n electrons per particle, this value will be raised by a factor $n^{\frac{1}{2}}$; but size range and shape remain unreasonable, since the curve is shifted to the right by the factor $n^{\frac{1}{2}}$. Reasonable assumptions as to the correction for the gravity effect (see Fig. 10) do not alter this picture appreciably.

The curve for $\beta=1$ is more reasonable as to size range and shape ($r\propto n^{\frac{1}{2}}$); but again the calculated $M(\alpha n^{\frac{1}{2}})$ is only 3.0 mg/m³, and this cannot be much increased by increasing n (number of electrons per particle at the easy-to-collect limit) if reasonable size range is to be maintained.

For the critical value of Case 2, $\beta=2$, a variation or spread in effective r of 3.9 is required to account for the spread in r^2/e values at 0.1 the maximum ordinate of Fig. 11. Even though there were no statistical fluctuations or discrete jumps in the relation $e=\alpha r^\beta$, departures from spherical shape would give a spread in r.

[20] E. W. B. Gill, Nature **162**, 568 (1948). See also V. E. Whitman, Phys. Rev. **28**, 1287 (1926).

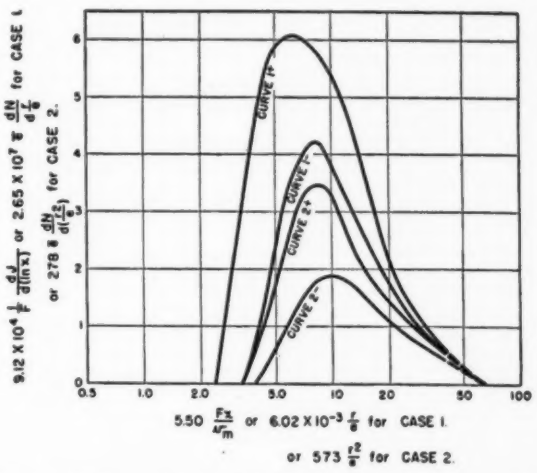


FIG. 11. Charge-weighted distributions for the silica aerosol data of Table I. Curves 1 are for samples taken directly from the exposure chamber; Curves 2 for samples passed through a 20-liter settling chamber. The + and − signs denote positive and negative charges, respectively. As indicated, the abscissas may be expressed in terms of charge and radius by the relation $r/e=(1/4\pi\mu Y)\times(Fx/v_m)$ for Case 1 (see Fig. 4), or by $r^2/e=[(A+B)L/4\pi\mu Y]\times(Fx/v_m)$ for Case 2. Similarly, the ordinates may be expressed in terms of number, charge, and radius by Eq. (8) of text for Case 1, or by Eq. (9) for Case 2. Aggregation and settling-out because of diffusion and gravity in the settling chamber cause the decrease in height and shift to the right of curves 2 over curves 1. The displacement of the − curves to the right of the corresponding + curves may indicate a tendency of positive charges to be associated with slightly smaller particles.

Although valid for Case 2 only, the above calculations result in some values of r larger than those covered by Case 2. If for Case 1 we assume $\beta=0$ and $n=10$ electrons per particle, the distribution of r (values αn) is not unreasonable, and $M(\alpha n^2)$ is 11 mg/m³. For $\beta=2$ and $n=1$ for the smallest particles falling at the right, the r values tend to be much too large ($r\propto n$), but $M(\alpha n^2)$ is 17 g/m³. Since the smallest particles certainly do not come under Case 1, these calculations for Case 1 are of interest only in that they show that the inclusion of large particles covered by Case 1 will tend to raise the calculated value of M, which is proportional to n^2 rather than $n^{\frac{1}{2}}$, as in Case 1. It is also pertinent, in this connection, that a fraction of the particles may be neutral and thus fail to contribute to our calculation of M, and that a few large particles deficient in charge will detract from our calculated value of M (or rather increase the actual value of M) in much greater proportion than their number.

From such considerations it is apparent that the charge-mass relations in our experiment cannot be described in terms of Case 1 or Case 2 alone, or possibly of a single value of α and of β. The magnitudes are such, however, as to admit the following reasonable possibilities:

(1) There is a large fraction of particles having a charge roughly proportional to surface area.

(2) Smaller particles tend to have unit electronic charge (more conclusive evidence is presented in reference 4).

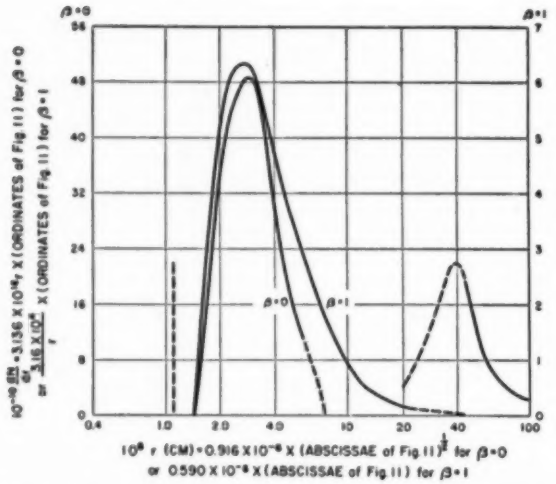


FIG. 12. Distributions of particle radii of Fig. 11, calculated from Eq. (11) on assumption that the particle charge $e=\alpha r^{\beta}$, that the smallest particles carry unit electronic charge, and that the particle radii are small compared to 640 angstroms (Case 2). The curve on the right is representative of those determined for similar dust particles by Dautrebande *et al.* (reference 9) with an optical microscope. Because the limit of resolution of the microscope is of the same order of magnitude, this curve can be used only as a rough indication of the upper limit of particle size, and is shown here probably on much too large a scale, compared to the β-curves (note difference in scale of the latter). The broken vertical line marks the mean radius reported in reference 9 as determined by electron micrographs of samples collected in a similar manner as those giving the optical microscope curve. Strict comparison with the β-curves is not justified because of differing circumstances (compare also Fig. 3, reference 4). Using the β-curves, aerosol mass concentrations for comparison with those determined experimentally with the electrostatic precipitator may be found by taking the area under curves constructed from these by multiplying ordinates by $4/3\pi\rho r^3$, ρ being particle density. A reassuring check on the accuracy of calculation and construction of these curves may be obtained by comparing the aerosol charge density determined by measuring the area under a curve of $\alpha r^{\beta}(dN/dr)$ *vs* r with that determined from the area under the $dJ/d(\ln x)$ *vs* x curve. Both methods give 6.7×10^{-4} esu/cm³ for the aerosol of Fig. 11. Because the calculated distributions extend beyond 640 angstroms, our assumption of Case 2 for *all* particles is not justified.

(3) An appreciable number of particles are larger than 640 angstroms, and a good portion of these larger particles, due probably to aggregation, are deficient in charge.

To investigate the charge-mass relation further, a cyclone-type apparatus was used to partially separate large from small particles. The aerosol entered tangentially at the top (diameter 2 inches) of an inverted, truncated cone of brass, and spiraled (over a vertical distance of 2 inches) to the bottom (diameter 1 inch), where smaller particles were removed from the center, and larger ones from the outer edge by two tubes through which equal air flows were maintained. Entrance nozzles of 0.058 and 0.080 inch diameter, and aerosol flows of 0.5 to 10 l/min (in each outlet tube) were used. No tests were made of separation efficiency; however, at the higher velocities, considerable silica was deposited in the cyclone chamber, and the charge collected was reduced by orders of magnitude.

From the discussion of Section 4, it appears that if $\beta>$ critical value (1 for Case 1, 2 for Case 2), then large particles will tend to be collected in the "easy-to-collect" region of low voltage, and small particles in the hard-to-collect region of high voltage. Therefore, if partial separation of the large from the small particles is effected, the current-voltage curve for the large particles should be greater in the low voltage region and less in the high voltage region relative to the curve for the small particles. Stated concisely, if the ratio R of large particle current to small particle current is greater in the easy-to-collect region than in the hard-to-collect region, $\beta>$ critical value, and vice versa.

The majority of test runs made gave practically the same value of R in the low and high voltage regions (taking gravity into account), with greater currents (up to a factor of 2) associated with the large particles. However, a few runs showed a tendency for R to be greater in the low voltage region ($\beta<$ critical value). A trend for a decrease in R in the low voltage region, indicating a decrease in effective β, was found when the aerosol was passed through a settling chamber before passing through the cyclone.

Thus, the data would seem to indicate that for aerosols taken directly from the chamber, the effective β lies near the critical value but decreases somewhat after passing through a settling chamber. The latter effect may be a result of aggregation (the effects of gravity and diffusion possibly would make for a more uniform size distribution and less efficient separation).

In view of the tendency of larger particles to be removed by the cyclone, it may well be that we are concerned here chiefly with particles covered by Case 2; in addition, more of the smallest particles may have been lost by diffusion in the longer path traveled through small diameter tubing. It is, therefore, reasonable to conclude from our results that among the particles which succeed in passing through the cyclone, there is an effective value of β of roughly 2. Perhaps the runs suggesting β to be somewhat greater than the critical value are an indication that some larger particles in the transition region between Case 2 and Case 1 are included,[21] since this condition would not require β to exceed 2 in these runs as it would have to otherwise.

If we assume that large particles deficient in charge and/or small particles with excess charge get through to the condenser, a value of $\beta>$ the critical value is necessary to explain the observed constancy of R. Such a value seems improbable for Case 2. However, a value of β between 1 and 2 could give the observed constancy of R if a portion of the particles of the large particle sample (this portion contributing most of the measured current) were of Class 1. The large effect of gravity on the easy-to-collect region of the curves of Fig. 10 argues against a low value of β, for gravity affects chiefly the large particles, and a low value of β

[21] This would be in accord with the fact that no such runs were found for the higher velocities which would more effectively remove transition or Case 1 (large) particles.

would give only small particles in this easy-to-collect region.

Thus, this experiment also is consistent with the hypothesis that a good portion of the particles have a charge which is proportional to the second (or slightly less) power of radius. It follows, from Sec. 4, that for such particles in Case 2, the shift of the charge-weighted distribution curves to larger abscissas (x or r^2/e) on passage through a settling chamber is occasioned almost entirely by aggregation, since gravity and diffusion shifts, besides being in opposite directions, are nonexistent for the critical value of β.

(b) Water Aerosols

Experiments concerned with dew points and light absorption[4] have shown that, when operating under equilibrium conditions at room temperature, a generator of the type we have used forms a water aerosol in a saturated or nearly saturated atmosphere at a temperature about 12°C below room temperature. As the aerosol travels down a glass tube, its temperature approaches room temperature, and evaporation maintains saturation until practically all the water has evaporated from the droplets, leaving extremely small, pseudo-solid residues.

Figure 13 shows charge-weighted distribution curves (plotted on a logarithmic scale in x) taken at three positions along such a tube: (1) near the generator (Curve 1); (2) further along, where particles are still liquid and the relative humidity practically 100 percent (Curve 2); and (3) still further, where little or no evaporation takes place, and the relative humidity is decreased (Curve 3). These are the curves resulting from the data presented earlier in Fig. 3. Flow in the 17-mm (i.d.) tube was probably turbulent, since its average velocity was 550 cm/sec, while the critical velocity for turbulence is 190 cm/sec. Flow in the condenser, however, was probably laminar, having an average velocity of 340 cm/sec, with 360 cm/sec the critical value for turbulence.

For these curves, the measured currents were nearly three orders of magnitude greater than those for samples taken from an exposure chamber because here, the aerosol flow is about 75 l/min (as compared to one or two) and is not diluted by additional air flow. Here, too, there has been comparatively little time for charge neutralization, since all measurements were made on the aerosol within the order of a second after its formation. The area under curve 3 corresponds to a charge density $\rho = 2.9 \times 10^{-2}$ esu/cm³.

Curve 1 extends very far in the direction of large x; in fact, its current-voltage characteristic is still rising at 900 volts. Hence, a large number of relatively large (hard-to-collect) particles[22] are present, and the size distribution is quite broad, even though peaked at the small (easy-to-collect) end. Because evaporation is practically complete for the particles of curve 3, these particles may be termed residual particles; they show the narrowest range in x.

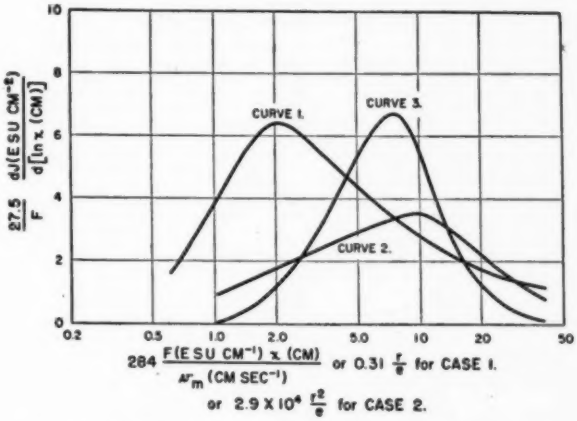


FIG. 13. Charge-weighted distributions for distilled water aerosols at increasing distances along the generator delivery tube. These curves are calculated from the original current-voltage data of Fig. 3. The area under the curves, when drawn with abscissas on a linear scale, is $A = 7.82 \times 10^3 (Jx/v_m)$. The aerosol charge density is $\rho = 4.03 \times 10^{-4}$A. Continuous evaporation along the tube maintains a reservoir of extremely small particles. From curve 1 to curve 3, aggregation, diffusion, and a dwindling source of these particles shifts the left-hand portion of the curves to the right, while evaporation shifts the right-hand portion to the left. The result is a narrowing of the distribution, which proceeds at a more rapid rate as the average size of the particles becomes smaller.

A large number of extremely small particles are present in all curves. Thus, the left-hand limit of curve 3 gives a particle radius of 13 angstroms on the assumption of unit electronic charge for Case 2, and curves 1 and 2 have even smaller particles. This is reasonable in view of the fact that evaporation is continuously producing residues of such size (from distilled water) immediately between the condenser plates, where they can be collected before being lost by diffusion or being enlarged by aggregation. That there should be so many such particles in curve 1 need not be surprising, since already at this position along the tube the temperature rise was one-half the value which gave complete evaporation between the positions of curves 2 and 3.

The left-hand portions of these curves show that aggregation (mainly of smaller to larger particles[4]), diffusion, and a depleting source of these very small particles were effective in producing a continuous shift to the right. The right-hand portions show that evaporation was effective in producing a continuous shift to the left corresponding to the gradual elimination of large liquid particles. The positions of the curve maxima are thus determined by this squeeze from either side. Probably, the reversal of the direction in which the maximum moved, between curves 2 and 3, is occasioned by the much more rapid shrinking in size of the smaller particles (relative to curve 1), due to evaporation.[23]

[22] That the hard-to-collect particles should be large assumes an effective $\beta < 2$ for Case 2, and $\beta < 1$ for Case 1. Evidence for this is given in reference 4.

[23] I. Langmuir, Phys. Rev. 12, 368 (1918).

7. SUMMARY AND CONCLUSIONS

We have presented an analysis of the current-voltage (dc) characteristic of a parallel-plate condenser through which a continuous flow of charged aerosol particles is maintained. Experimental conditions, involving apparatus dimensions, rates of flow, particle sizes, etc., under which a charge-weighted distribution of radius-charge ratio may be obtained for the aerosol particles are specified. Fortunately, the latitude in these conditions proves to be large. For particles small compared with the mean free path of air molecules, the second power of radius, rather than the first, appears in the radius-charge ratio.

Since the electronic charge establishes a lower limit for particle charge, it was possible to calculate size distributions on various assumptions concerning the relation between charge and radius. From the reasonableness of these distributions, and their ability to check the experimentally determined mass concentrations, many possible assumptions were excluded, and a reasonable charge-radius relationship for a silica aerosol under the conditions of our experiments was proposed. The nature of this relationship was further checked by measurements on samples where partial separation of large and small particles was achieved.

For a silica aerosol, the charge-weighted distribution of radius-charge ratio was shown to be sensitive to the effects of "settling out" by diffusion and gravity, and to aggregation. For a water aerosol, the additional effect of evaporation was demonstrated.

Although the experiments carried out have been chiefly exploratory, and were designed to test the ideas summarized briefly above, the following interesting observations have resulted:

(1) Silica aerosols, produced as described in Sec. 2, and existing in an exposure chamber under conditions described in that section, contain many particles which are larger than the mean free path of air molecules (640 angstroms). However, the great majority of particles are smaller than this (most probable radius between 200 and 300 angstroms), some probably approaching molecular dimensions.

(2) There seems to be a tendency for the greater number of negative particles to be associated with a slightly larger particle size than is the case for the greater number of positive particles in silica aerosols.

(3) For a large proportion of silica particles, the particle charge appears to be proportional to the square of the radius. Deviations (other than statistical fluctuation) from such a relationship result for very small particles when it would require that their charge be less than a single electron, and for larger particles when they have been formed by aggregation with consequent charge neutralization.

(4) Data for aerosols of distilled water passing down a 17-mm (i.d.) tube show that, as the temperature is raised over the 4°C or so necessary to evaporate the water (leaving residue particles), a reservoir of a large number of small particles is maintained by the competing actions of evaporation in producing these particles, and of diffusion and aggregation in removing or enlarging them.

It will be noted that in the application of our method we have been concerned only with the more complicated of the situations outlined in the Introduction. For such situations, no completely rigorous description can be obtained, and to deal with them at all, it is necessary to exercise critical judgment in weighing whatever information can be obtained. The method we have described is of assistance in furnishing additional information and in presenting it in a form which can be handled with some degree of objectivity. In a separate article,[4] we give more complete examples of this type of procedure, particularly in connection with aggregation phenomena and with comparisons between aerosols.

It is apparent that a more complete description can be obtained by the method in the case of charged particles[24] of known constant size. It is here that the method seems to offer greatest promise. By assisting the development of methods of controlling particle charge, and by establishing such aspects of the problem as the effect of charge on the behavior of particle clouds, the method should contribute to a general understanding of the nature and behavior of aerosols.

8. ACKNOWLEDGMENT

The authors wish to thank Dr. L. Dautrebande, at whose request this investigation was undertaken, for the use of his aerosol generating equipment. Mr. W. C. Alford assisted materially in the early stages of the investigation, as did Mr. Lawrence Showkeir. We particularly wish to express our appreciation to Dr. J. M. Dalla Valle and Mr. T. F. Hatch for their helpful criticisms and suggestions regarding the manuscript.

[24] Effects of uncharged particles are considered in reference 4.

New Appointment

Dr. Robert F. Bacher has been named chairman of the Committee on Atomic Energy in the Department of Defense Research and Development Board. Dr. Bacher is head of the Physics Department at the California Institute of Technology.

Purpose of the Committee on Atomic Energy, composed of four civilian and six military members, is to assist the Research and Development Board in the atomic energy aspects of the military research and development program, and to coordinate the research and development activities of the Department of Defense with those of the Atomic Energy Commission.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Some Quantitative Data Bearing on the Scabbing of Metals under Explosive Attack

JOHN S. RINEHART
U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received August 8, 1950)

Fracturing, or scabbing, of a material near a free surface as the result of a transient compressional stress wave of high intensity impinging on that surface has been observed for many years; however, little quantitative data that relate the fracture to the nature of the stress wave and the physical properties of the material seem to exist. The phenomenon has been investigated for five metals, 1020 steel, 4130 steel, 24S-T4 aluminum alloy, brass, and copper, by using an explosive charge to induce a high intensity stress wave in the metal. The distribution of pressure within the wave was determined by a modified Hopkinson-bar type of experiment.

Scabbing has been found to be governed principally by the spatial distribution of pressure within the wave and a critical normal fracture stress σ_c that is characteristic of the material and perhaps the state of stress. Numerical values of σ_c were obtained for each of the five metals.

INTRODUCTION

A COMMON and, in many cases, striking effect produced by a transient compressional stress wave of high intensity is scabbing. Scabbing, sometimes called spalling, is the fracturing of a material near one of its free surfaces which is relatively far removed from the area of application of the pressure (or stress) impulse. Such fracturing occurs often in ordnance work. Under some conditions a small explosive charge, detonated on one side of a heavy steel plate, will cause a small section to fly off the back side without producing any apparent damage to the interior of the plate. Viewed qualitatively, the metal fractures because its strength is exceeded as a result of the reflection of the compressional stress wave as a tension wave. The scabbing of a concrete wall when struck by a projectile, or the breakup of a high speed missile on impact are other well-known and typical examples. Little quantitative data concerning the effect seem to exist, although scabbing has been reasonably well understood qualitatively for many years. Some quantitative information has been reported on the fracturing of concrete piles during driving, another example of the same type of effect.[1,2]

The investigation described here was carried out for the purpose of determining some of the factors that control the nature of the scabs produced in certain metals by explosive charges. An important part of the investigation was the experimental determination of the intensities and durations of the pressures (or stresses) generated within the metals by the explosions.

MECHANICS OF SCABBING

Geometrical considerations indicate that, if a plate scabs as the result of reflection of a high intensity transient compressional stress wave at a free surface, the thickness of the scab ought to be governed principally by two factors: (1) the shape of the stress wave and (2) a critical normal fracture stress that is characteristic of the material acted upon. The latter may be much influenced by the conditions of loading and the state of stress in the body.

Consider a laterally infinite slab of material of finite thickness and assume a perfectly elastic plate material. Suppose that the left face is subjected to a stress that varies with time. Further, suppose that the pressure rises suddenly and then begins to decay. The stress condition inside the plate, shortly after the impulse has struck, will be that shown in Fig. 1(A). When the stress wave strikes the free (right) boundary of the plate, it will be reflected as a tension wave. Interference results between the incident compression wave and the reflected tension wave. The resultant stress distribution in the plate at this time is shown in Fig. 1(B). The tension AB is seen to increase as the reflected wave moves to the left. At some particular point, the metal may no longer be able to support the tension. It then fractures along the line MN in the figure, and a scab of thickness δ flies off. The tension AB, under which the material fractures, will be referred to here as the critical normal fracture stress σ_c of the material. Scabbing should always occur, therefore, under the conditions considered here, whenever the maximum value of the stress within the wave exceeds σ_c and should never occur when this is not the case. The thickness of the scab ought to be equal to one-half the distance within the wave that corresponds to a decrease in stress equal to σ_c.

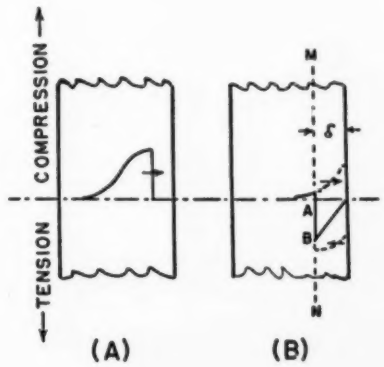


FIG. 1. Manner in which incident and reflected waves interfere.

[1] E. N. Fox, Engineering **134**, 263 (1932).
[2] Glanville, Grime, and Davies, J. Inst. C. E. **1**, 150 (1935).

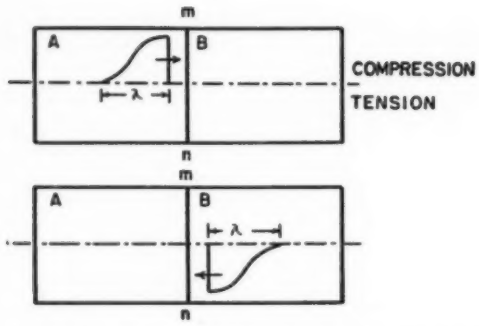


FIG. 2. Principle of the Hopkinson method. (Top) Before stress wave has reached interface. (Bottom) Reflected stress wave.

Experimentally, the procedure has been to determine the stress wave, measure the scab thickness, and then deduce σ_c from these data.

GENERAL METHOD

The method that has been employed to obtain stress information is somewhat similar to that used first by Hopkinson and later by others.[3–5] Essentially it consists of the experimental determination of successive increments of momentums, i.e., areas under the stress-time curve and, from these, construction of the complete curve.

An extensive treatment of the basic ideas of the method was given by Hopkinson. It seems worthwhile, however, to review them here for the sake of completeness. Consider two plates A and B, of similar material and unit cross section, assumed joined along the line mn (Fig. 2). The joint has no strength in tension but will transmit a compressional stress wave without change. Suppose, now, that a transient compressional stress wave is moving toward the right through A. It will enter unchanged into B but, on reaching the free (right) surface of B, will be reflected as a tension wave, the condition shown in the lower drawing of Fig. 2. When the thickness of B is greater than one-half the wavelength λ of the stress wave, it is apparent that, when the return wave reaches the interface between A and B, plate B will fly off to the right with all of the momentum trapped in it. Plate A will remain at rest.

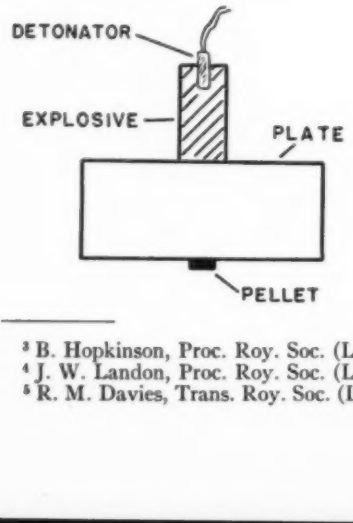


FIG. 3. Cutaway view of charge-plate-pellet arrangement.

The momentum possessed by B will be the total area under the stress-time curve.

When the thickness of B is made less than ½λ, only a part of the momentum will be trapped in B; the remainder will not get out of A, provided the wave has a sharp front. The amount M_B trapped in B will be given by

$$M_B = \int_0^T \sigma(t)dt,$$

where T is the length of time it takes the wave to travel across B and back again, and $\sigma(t)$ is the stress as a function of time. Experimentally, B is made initially very thin and then its thickness is increased, by steps, up to and beyond the point where its momentum reaches a constant value. By measuring the momentum of the plate B in each case the stress-time curve can be deduced.

Hopkinson, in his original experiments, used for A and B two cylinders suspended as ballistic pendulums and measured the momentum imparted to each as a result of a bullet striking A. By varying the length of B, he was able to obtain the retarding force on the bullet as a function of time.

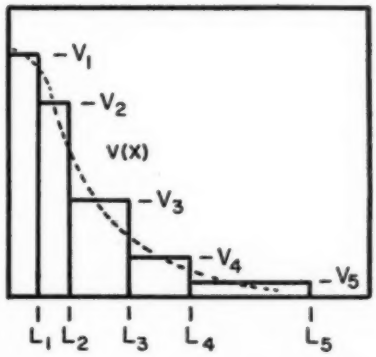


FIG. 4. Method of calculating particle velocity distribution.

The experimental arrangement in the present case was considerably different. A small cylinder of explosive was placed on one face of a heavy metal plate and a small pellet of the same material placed on the opposite face in line with the charge. A cutaway section of the arrangement is shown in Fig. 3. The plates were about 5 inches in diameter and ranged in thickness from 1½ in. to 3 in. Usually, five different thicknesses of pellets were used. Nominal thicknesses were $\frac{1}{16}$ in., $\frac{1}{8}$ in., $\frac{1}{4}$ in., $\frac{3}{8}$ in., and $\frac{3}{4}$ in.

Specifically, this was the experimental procedure. Five plates of the same thickness and five pellets that had the respective nominal thicknesses given above were selected. A charge and one of the pellets were affixed to the plate, as shown in Fig. 3. The charge was detonated and the velocity of the pellet measured. This procedure was repeated with pellets of different lengths. A new charge and plate were used each time.

The measured velocities of the pellets yielded five increments of the curve shown in Fig. 4. This curve

[3] B. Hopkinson, Proc. Roy. Soc. (London) **A213**, 437 (1914).
[4] J. W. Landon, Proc. Roy. Soc. (London) **A103**, 622 (1923).
[5] R. M. Davies, Trans. Roy. Soc. (London) **A240**, 375 (1948).

describes the change in velocity of the surface of the plate as the wave advances against it. The ordinates of the respective increments are designated by $v_1, v_2, \cdots$. The ordinate v_i of a particular increment is seen to be

$$v_i = \frac{V_i L_i - V_{i-1} L_{i-1}}{(L_i - L_{i-1})} \quad i = 1, 2, 3, 4, \text{ and } 5,$$

where $V_1, V_2 \cdots$ are the measured velocities of the pellets. The quantity $(L_i - L_{i-1})$ is the width of the increment and is equal to *twice* the difference in thickness between two successive pellets. Nominal values for $L_1, L_2, \cdots$ are, therefore, just twice the thicknesses of the pellets used in these tests, i.e., $\frac{1}{8}$ in., $\frac{1}{4}$ in., $\frac{1}{2}$ in., $\frac{3}{4}$ in., and $1\frac{1}{2}$ in.

A curve of the type given in Fig. 4 can be easily transformed to a stress-time curve, provided that all parts of the wave are assumed traveling at the same velocity. For each point in the wave, the pressure (or stress) σ at that point will be given by

$$\sigma = \rho c v,$$

where v is the particle velocity, ρ is the density of the material, and c is the velocity of the wave. The general velocity of sound for waves of dilation seems to be most appropriate here. This velocity is given by

$$c = (3K(1-\nu)/\rho(1+\nu))^{\frac{1}{2}},$$

where K is the bulk modulus, and ν is poisson's ratio. Only approximate values of c were used. Specifically, these were 217,000 in./sec for 1020 steel and 4130 steel, 151,000 in./sec for brass, 154,000 in./sec for copper, and 224,000 in./sec for 24*S*-*T*4 aluminum alloy. No account was taken of any variation of K or ν with pressure.

TESTS AND MATERIALS

The above method has been used to determine the distributions of pressure. Curves were obtained from plates of four thicknesses, $1\frac{1}{2}$ in., 2 in., $2\frac{1}{2}$ in., and 3 in. for each of the five metals mentioned above.

In each case, the charge was a cylinder 1 in. in diameter and 2 in. long. It was formed by packing Composition *C*3 explosive into a cardboard container. The charge was detonated with an Engineer's Special cap, inserted approximately $\frac{1}{2}$ in. into the explosive.

The plates ranged in diameter from 5 in. to 6 in., i.e., they were essentially infinite laterally. Both surfaces of each plate were ground so that, on one face, the explosive would be in intimate contact with the surface and, on the other, the joint between the pellet and the plate would be good. The corresponding surface of the pellet was also ground. Rubber cement was used to affix the pellet to the plate. The explosive was well bonded to the surface of the plate because of the former's oil base. The pellets, in all cases, were $\frac{1}{2}$ in. in diameter and were made of the same material as the plate.

TABLE I. Description of materials.

Plate material	Composition Percent	Constituent	Heat treatment
Copper	99.81	Cu	Annealed at 850°F for 2 hr. Cooled slowly.
Brass	38.62	Zn	Annealed at 850°F for 2 hr. Air cooled.
	0.75	Sn	
	60.63	Cu	
4130 Steel	0.32	C	Annealed at 1575°F for 10 hr. Cooled slowly.
	0.60	Mn	
	0.29	Si	
	0.017	P	
	0.012	S	
	0.93	Cr	
	0.23	Mo	
1020 Steel	0.23	C	Annealed at 1575°F for 10 hr. Cooled slowly.
	0.94	Mn	
	0.22	Si	
	0.024	P	
	0.134	S	
	—	Cr	
	—	Mo	
24*S*-*T*4 Aluminum Alloy	4.40	Cu	Used as received in *T*4 condition.
	1.29	Mg	
	0.27	Si	
	0.74	Mn	

It was feared, at the beginning of the investigation, that variability in the action of the explosive from plate to plate might introduce large inconsistencies in the results. This fear was not substantiated. Successive firings generally gave results which agreed within a few percent.

Table I presents a list of the materials, their composition and any pre-test heat treatment. The velocity of each pellet was measured over about a 3-ft interval by means of special multiflash photography. Usually, two or more tests were run on each plate-pellet combination.

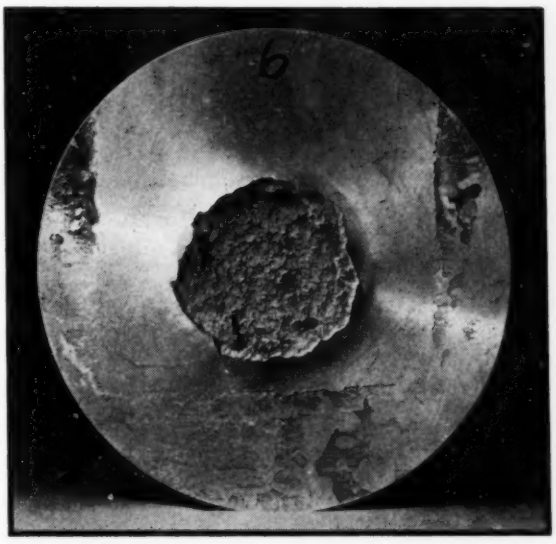


FIG. 5. Photograph of scabbed aluminum alloy plate.

TABLE II. Summary of scabbing data.

Material	Thickness of plate (in.)	Scab thickness δ (in.)
24S-T4	1½	0.13
	2	NS*
	2½	NS
	3	NS
Copper	1½	0.13
	2	0.13
	2½	NS
	3	NS
1020 Steel	1½	0.14
	2	0.17
	2½	0.25†
	3	NS
Brass	1½	0.09†
	2	NS
	2½	NS
	3	NS
4130 Steel	1½	0.14
	2	NS
	2½	NS
	3	NS

* No scabbing observed.
† Only faint crack observable.

OBSERVED SCABBING

Generally, the thinner plates scabbed and the thicker ones did not. Sometimes a small section was actually torn loose from the plate and projected forward with considerable velocity; at other times, the plate had to be sectioned in order to observe the fracture. A photograph of a typical scabbed plate is reproduced in Fig. 5. The scab does not extend over the whole surface of the plate. This is to be expected since the stress wave set up by the charge must be divergent, so that the intensity of the stress becomes less at greater distances from the axis. Over the area of the plate covered by the pellet, a ½-in. diameter circle, the wave has been found to be substantially plane.

Complete scabbing data are summarized in Table II. Plates that carried no pellets were fired in order to obtain specimens for scab thickness determinations. This was necessary since an affixed pellet is essentially an extension of the plate and would be expected to affect the thickness of the scab. The fractures were brittle and the surfaces of the scabs were fairly rough. There was an uncertainty of about ± 0.01 in. in the measurement of the thickness of any one scab.

PRESSURE-TIME CURVES

Significant pressure-time curves that have been determined by the above method are plotted in Figs. 6A–6E. The curves have been plotted as histograms, since this type of plot represents most accurately the experimental data. The successive rectangles correspond to velocity measurements made on $\frac{1}{16}$-in., $\frac{1}{8}$-in., $\frac{1}{4}$-in., and $\frac{3}{8}$-in. thick pellets, respectively. The height of the rectangle corresponding to the $\frac{3}{4}$-in. pellet was found to be nearly zero in all cases. Although four histograms were obtained for each metal, only those directly related to the present discussion of scabbing have been included. The dotted smooth curve, sketched on each histogram, is the author's guess as to the probable shape of the true pressure curve.

Each curve describes the pressure at the free surface of the plate as the wave impinges on it. This pressure is substantially, but not exactly, the same as the spatial distribution of pressure within the plate close to its surface. The two differ because energy is extracted from the wave as it passes through the plate and hence the shape of the wave changes progressively.

The progressive change in the shape of the wave as it passes through the metal is readily apparent from the curves. In every case, the high intensity part of the curve decays most rapidly. This is to be expected. The low intensity part of the curve remains substantially the same in the case of copper and brass, although it shows a slight attenuation. The latter absorption is probably caused by divergence of the stress wave. As pointed out above, the explosive charge does not set up a perfectly plane wave in the metal plate. In the case of the steels, the low intensity part of the wave seems to build up as it travels through the plate. It appears that in some way energy is fed from the front of the wave into its rear. This transfer may be related to different rates of propagation of the different parts of the wave.

CRITICAL NORMAL FRACTURE STRESS

The critical normal fracture stress, σ_c, for each of the metals can be found by using the curves of Fig. 6 and the data on scabbing given in Table II. The following procedure was used. The thickness of the scab was multiplied by two and divided by the velocity of sound in the material. This yielded a time. The stress that corresponds to the time found in this way, when subtracted from the maximum ordinate of the stress-time curve, gave σ_c. No value for σ_c can, of course, be calculated for those plates that did not scab. The fact that a plate did not scab must mean that σ_c is greater than the maximum pressure that existed in the pressure wave.

The two copper plates that scabbed yielded values for σ_c of 430,000 lb/in.² for the 1½ in. plate and 400,000 lb/in.² for the 2 in. plate. The 2½ in. plate did not scab; however, the maximum pressure in this case was found to be 280,000 lb/in.², i.e., a value lower than σ_c.

Only one plate, the thinnest, of brass, 4130 steel, and 24S-T4 aluminum alloy scabbed, so that only a single value of σ_c was obtained in each case. These values were 310,000 lb/in.², 440,000 lb/in.², and 140,000 lb/in.², respectively.

Two significantly different values of σ_c were found for 1020 steel. The 2 in. plate gave a value of 230,000 lb/in.², while the 2½ in. plate gave 130,000 lb/in.² The highest pressure measured at the surface of the 3 in. plate was 135,000 lb/in.² This plate did not scab.

It is possible that the state of stress that exists in

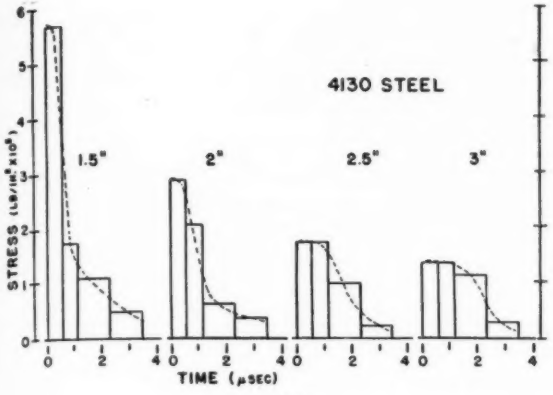


FIG. 6(A)

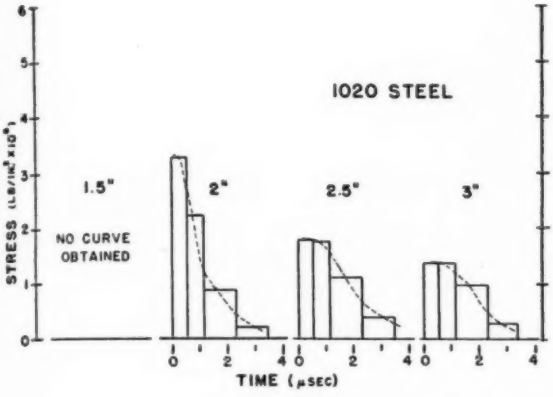


FIG. 6(B)

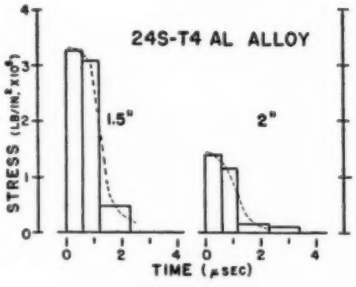


FIG. 6(C)

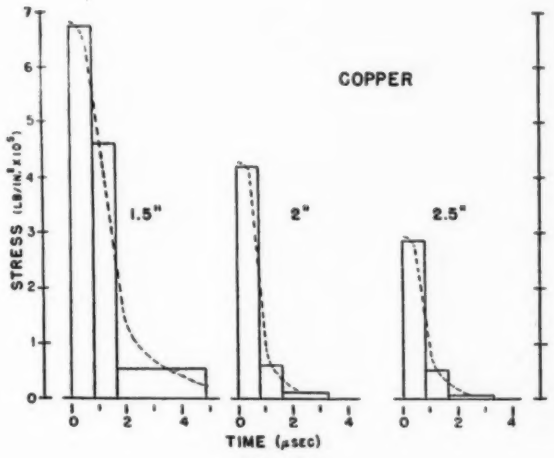


FIG. 6(D)

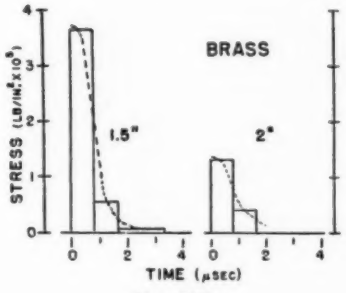


FIG. 6(E)

FIG. 6(A)–6(E). Pressure-time curves for: (A) annealed 4130 steel, (B) annealed 1020 steel, (C) 24*S*-*T*4 aluminum alloy, (D) annealed copper, and (E) annealed brass.

the region of the fracture at the time the fracture occurs may affect the value of σ_c. The ambient pressure at the time the scab was formed in the 2 in. 1020 steel plate was about 90,000 lb/in.²; on the other hand, the ambient pressure was only 45,000 lb/in.² in the 2½ in. plate. Such an increase in normal stress to fracture is in qualitative agreement with Bridgman's[6] observations that the resistance to cleavage fracture is increased by application of hydrostatic pressure.

No similar pronounced effect seemed to exist in the case of copper. The ambient pressures at which scabbing took place differed by a factor of ten. They were 250,000 lb/in.² for the 1½ in. plate and 25,000 lb/in.² for the 2 in. plate. Respective values of σ_c differed by only a few percent.

The data were insufficient to establish whether such an effect exists for 24*S*-*T*4 aluminum alloy, brass, or 4130 steel, since only one value of σ_c was found in each case.

The critical normal fracture stress of annealed 4130 steel was approximately twice that of annealed 1020 steel. On the other hand the stress waves decayed in the two materials in about the same manner. Such behavior suggests strongly that the plastic flow mechanism responsible for the decay is relatively unaffected by an alloying agent, whereas the resistance to fracture may be greatly increased.

In so far as the author knows, the data reported here are the only quantitative data available on the resistance to fracture under the high rates of loading and

[6] P. W. Bridgman, "Fracture and hydrostatic pressure," *Fracturing of Metals* (American Society for Metals, 1948), p. 246.

TABLE III. Comparison of data with Clark and Wood's critical impact velocities.

	Critical normal fracture stress σ_c (lb/in.²)	Associated critical impact velocity (ft/sec)	Clark and Wood's critical impact velocity (ft/sec)
24S-T4	140,000	202	290*
Copper	430,000	277	235*
	400,000	258	
Brass	310,000	216	†
1020 Steel	230,000	120	100
	130,000	68	
4130 Steel	440,000	235	†

* Calculated from static stress-strain curve.
† Not determined.

states of stress produced by explosive charges. It is of interest, however, to compare these results with theoretical studies of von Karman and Duwez,[7] and Clark and Wood's[8] experimental investigations of the behavior of one dimensional specimens under longitudinal impact. According to von Karman's theory, there is an impact velocity at which rupture will occur near the moving end of a specimen that has been subjected to impact tensile loading before any appreciable plastic strain takes place along the specimen. Above this velocity, designated as the critical impact velocity, the specimen will break with very little elongation. The critical velocity can be computed approximately from the static stress-strain diagram. Clark and Wood[8] have determined experimentally the critical impact velocity for a large number of materials. Their experimental results agreed closely with their theoretically computed values.

[7] T. von Karman and P. Duwez, J. Appl. Phys. **21**, 987 (1950).
[8] D. S. Clark and D. S. Wood, Trans. Am. Soc. Metals **XLII**, 45 (1950).

It was assumed early in the present paper that a constant relationship exists between the pressure at a point in the pressure wave and the velocity of a particle at that point. The same relationship,

$$\sigma = \rho c v$$

can be used to associate a critical particle velocity with each critical normal fracture stress. Values of critical particle velocities obtained in this way are listed in Table III. Critical impact velocities found by Clark and Wood[8] are listed in the same table. The two velocities are seen to be about equal numerically. It seems not unlikely that, through further study, they can be shown to be physically equivalent.

CONCLUSIONS

The results that have been obtained suggest strongly that the physical processes involved in scabbing, postulated in the early part of the paper, are substantially correct. Scabbing has been found to be governed principally by the spatial distribution of pressure within the wave and by a critical normal fracture stress that is characteristic of the material under attack. There is some evidence that the latter may be influenced by the state of stress existing in the material at the time fracturing occurs.

ACKNOWLEDGMENT

In conclusion the author gratefully acknowledges the assistance of several Naval Ordnance Test Station personnel. Mr. C. C. Woolsey and Dr. W. O. Wetmore showed great interest in the work and made many helpful suggestions. Mr. Ernest Barkofsky provided the multi-flash photography equipment and Mr. Guy Throner, the explosive facilities. Mr. Clyde McCrary made many of the firings and reduced much of the data.

Awards and Honors

Five promising young scientists, including a woman mathematician, were named by Bell Telephone Laboratories to receive the 1951–52 Frank B. Jewett postdoctoral fellowships. The awards, designed to stimulate and further the work of researchers in the physical sciences, grant $3000 to the recipient and $1500 to the institution at which he chooses to do his research.

Winners are: Murray Gerstenhaber of the University of Chicago and New York City; Donald Roy Francis Cochran of Johns Hopkins University and San Francisco, California; Dr. Ilse Lisl Novak of the University of California and New York City; Stephen Prager of Cornell University, Ithaca, New York, and Donald Robert Yennie of Columbia University and Midland Park, New Jersey. Two of the winners are chemists, two are mathematicians, and one is a physicist. Dr. Novak was also among the award winners last year.

Grants for the fellowships were established in 1944 by the American Telephone and Telegraph Company, upon the retirement of the late Dr. Jewett as Vice-President in charge of Development and Research. Since that time 38 fellowships have been awarded.

The fellowships are designed to stimulate and assist research in the fundamental physical sciences and particularly to provide the holders with opportunities for individual growth and development as creative scientists. Jewett fellows have conducted research at most of the country's leading universities and institutes. Each recipient is free to select the institution at which he will do his research.

The fellowships are awarded on recommendation of a committee consisting of seven members of the technical staff of Bell Telephone Laboratories. Primary criteria are the demonstrated research ability of the applicant, the fundamental importance of the problem he proposes to attack, and the likelihood of his growth as a scientist. The awards are postdoctoral, and only scientists who have recently received doctor's degrees or who are about to receive them are normally considered.

The American Crystallographic Association paid tribute to the National Bureau of Standards, on the occasion of its Fiftieth Anniversary, for its many contributions to science and technology. A commemorative scroll was presented by Dr. R. W. G. Wyckoff, President of the ACA, at the Association's semiannual meeting in Washington, D. C., on February 15–17. More than 200 leaders in the science of crystallography from this country and abroad attended the meeting, which was held at the Bureau as part of its semicentennial celebration.

During the award ceremonies, Dr. Wyckoff noted the Bureau's 50 years of scientific service and presented the scroll. Accepting the scroll on behalf of the Bureau, Dr. E. U. Condon, NBS Director, stressed the Bureau's future plans for fundamental research in the physical sciences.

Permanent Magnet Lenses*

JOHN H. REISNER
RCA Victor Division, Radio Corporation of America, Camden, New Jersey
(Received August 21, 1950)

Permanent magnets have been successfully utilized to energize magnetic lens pole pieces for high resolution electron microscopy at direct magnifications of 7000 times with 30-kilovolt electrons, and 6000 times at 50 kilovolts. Stray magnetic fields, which have previously limited high magnification and high resolution, have been minimized by using two circuit gaps in parallel and by surrounding the assembly with a magnetic shield connected to the external pole pieces of the two gaps. Magnetomotive forces of 1400 gilberts have been obtained. A single magnet, magnets in bucking arrangement, and configurations involving magnets disposed either radially or parallel to the optical axis of the lens, have been successfully utilized. The effects of various magnetic circuit parameters on the design and operation of the lens system have been studied.

INTRODUCTION

THE permanent magnet as an energizing source for magnetic electron lenses is not new. The use of a permanent magnetic yoke for the comparatively coarse focusing of cathode-ray tubes has long been known. The advantages of permanent magnet lens energization are very appealing—excellent stability (beyond the ability of any regulator), no heating losses in energizing coils, no need for extensive current supplies and regulators—advantages which heretofore were limited to electrostatic lenses. Undoubtedly, the limitation which earlier prevented the permanent magnet lens application to the electron microscope was one of insufficient energization. With the development of new magnetic alloys during the past ten years, the use of permanent magnet lenses with the electron microscope was undertaken.

Three applications[1–3] are known where permanent magnet lenses have been substituted for energizing coils. In the case of B. v. Borries *et al.*,[1] and Zworykin and Hillier,[2] the substitution actually changed the magnetic circuit by replacing a low reluctance coil shield by a high reluctance energizing unit. The result was the production of powerful stray fields which had to be dissipated or shielded against. Such stray fields decrease the magnification of the final image and frequently cause distortion or aberration to hurt resolution in the final image.

Kinder and Pendzich[3] started with a double-yoke lens whose construction was such that the substitution of permanent magnet energization did not greatly alter conditions of stray flux. They provided heavy shielding about the optical axis to minimize the effect of this field.

The great difficulty in sufficiently shielding the electron beam in applications, such as in (Fig. 1), caused Ramberg† to approach the problem from a different standpoint. Opposing magnets were used to produce an assembly having sensibly equal potential ends with a side shielding to localize leakage flux from the magnets. The result is the elimination of serious stray end fields. This is not the first record of bucking magnets. Nicoll† shows magnets in opposition, but having no external shielding to connect equipotential components together and eliminate localized stray fields.

Of interest, but not significant at the present state of

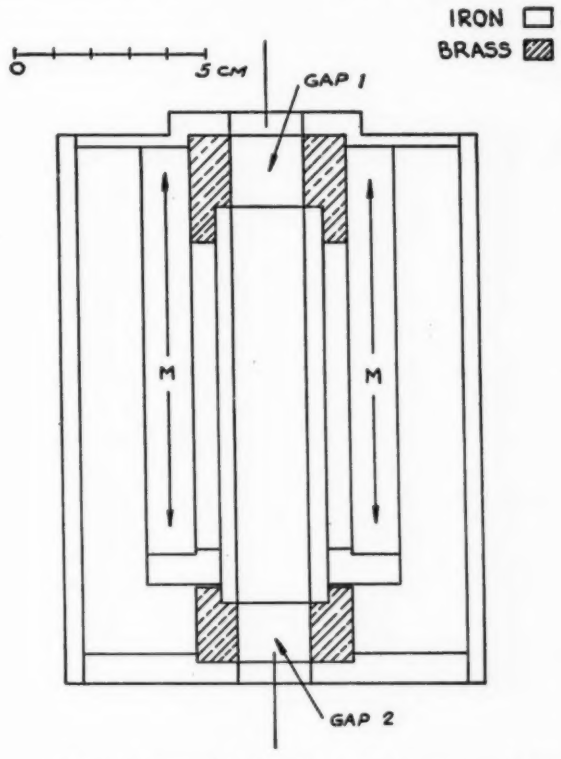


FIG. 1. An actual design of an axial arrangement of magnets in a two-gap lens system.

* A paper delivered before the Electron Microscope Society of America, October 6, 1949, by John H. Reisner. J. Appl. Phys. **29**, 68 (1950).

[1] Borries, Ruska, Krumm, and Muller, Naturwiss. **28**, 350 (1940).

[2] V. Zworykin, J. Hillier, J. Appl. Phys. **14**, 658 (1943).

[3] E. Kinder, and A. Pendzich, Jahrb. der A. E. G., Forschung **7**, 23 (1940).

† Patent literature is perhaps the only source of information concerning several important ideas pertaining to permanent magnetic lenses. Particularly important is the work of E. Ramberg, U. S. Patent 2,369,796. Of interest are British Patent 576,369, on permanent magnet pole pieces, and F. H. Nicoll, U. S. 2,200,039, on unshielded bucking magnets, cited in the text.

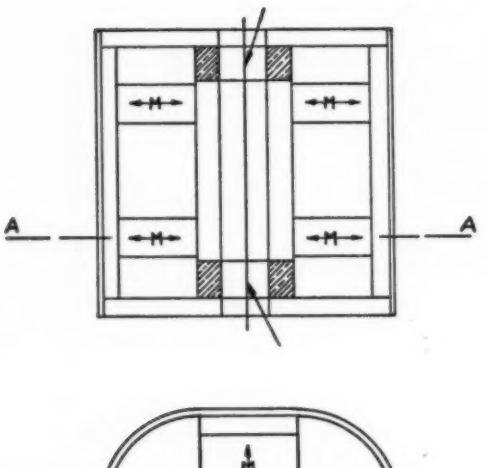


FIG. 2. An actual design of radial magnets in a two-gap lens system.

the magnet art, is the fabrication of the pole pieces from the magnet material. Several objections may be raised to such construction. The magnet materials do not lend themselves to the careful machining required in pole pieces, nor is their structure or magnetization sufficiently homogeneous. Furthermore, the magnetic potential drop along the inside of such a pole piece gives rise to a field supplementary to the fields in the gap. The action of such a secondary field is extremely undesirable for electron lenses.

BASIC PRINCIPLE

Stray magnetic fields in an electron optical system may be expected to give rise to distortions, deflections, and defocusing, much as a poorly figured optical surface might affect a light optical system. It is conceivable that leakage flux from the ends of a magnetic lens energizing unit may be sufficiently radially symmetrical merely to act as additional lenses, making it possible to produce focused optical images under certain conditions. However, such fields generally operate as high aberration lenses. Also, the inclusion of any ferrous material, such as screws, brackets, etc., in the space occupied by the field, would result in a destruction of the radial symmetry of the field, giving rise to a distorted image. In optical systems designed for the use with coils of an axially concentric winding, the simple substitution of permanent magnets can frequently give rise to stray fields which render the beam shaping function of the electron source inoperative.

Theoretically, it is possible to reduce the stray fields below any desired value by shielding methods. However, from a practical viewpoint, this has obvious mechanical disadvantages. Another method for avoiding stray fields in the optical axis is to direct the stray field away from the axis of the optical system. Both of the methods above have the operational hazard of magnetizing any magnetic material brought near to the apparatus, as for example, wrist watches and tools.

In designing a system free of external fields, the following principle is basic. *Wherever a high reluctance source of MMF is used to produce a radially symmetrical magnetic field, with the requirement that the axis of such a field not pass through the magnet material, it is necessary to provide two shunting high reluctance paths (gaps) for that source of MMF.*

Such a principle leads one to the conclusion that a double-pole piece lens system is the logical way to use both of the necessary high reluctance gaps. When the exterior pole pieces of each of the two gaps of such a system are connected by a low reluctance circuit, the difference in MMF between the ends of the assembly will be small, and, consequently, the stray field will be negligible. Actual embodiments of such an energizing means are shown in Figs. 1 and 2. Recent literature[4] shows a parallel development giving rise to excellent results.

Several interesting and valuable properties accrue from these designs. The gaps are energized in parallel but with fields oppositely directed so that optical rotation of the image, which always takes place in magnetic lenses, is in the opposite direction in the two gaps. The result is to reduce the rotational effect far below that of a single lens. Lenses have been built where the rotation was negligible. Since the rotational property is fre-

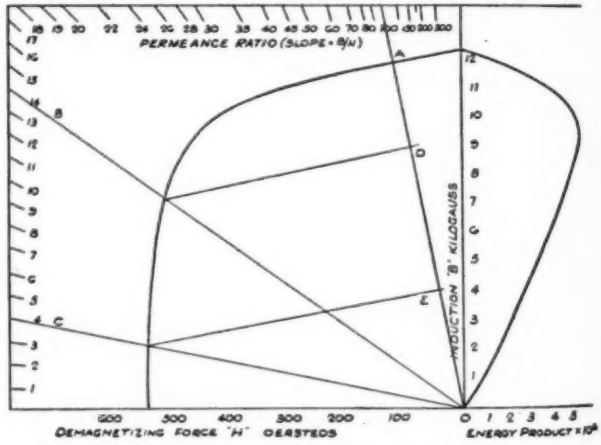


FIG. 3. *B* and *H* diagram for axially disposed magnets. Demagnetization curve, Alnico V.

[4] B. v. Borries, Kolloid-Z. 114, 164 (1949); *Die Ubermikroskopie* (Saenger, 1949), p. 104.

quently used as a criterion for alignment of optical components and lenses, the lack of rotational effect is not always desirable. However, in cases where alignment is not perfect and where focusing is accomplished by varying the high voltage which accelerates the electron beam, such magnetically opposed lenses do reduce or stop the rotational effect which would otherwise occur as the voltage was varied. The stable image resulting is an aid to the operator trying to focus an image.

DESIGN CONSIDERATIONS

The feature of low stray field is not the only requirement for a successful permanent magnet lens design. Strength, ability to remove pole pieces without demagnetizing the magnets, and means of energizing the magnets are all problems of importance.

Several types of lens assemblies are possible. In Figure 1 is shown a cross section of an assembly using four axially disposed rectangular magnets. Measurements of flux on such an assembly give magnetomotive forces of 1300 gilberts in gap 1 and 900 gilberts in gap 2. The position of operation on the B and H curve is shown in Fig. 3, corresponding to a permeance ratio of 100. Operation of the magnets occurs at an energy product of 30 percent of maximum. From a design point of view, this represents poor utilization of magnetic material. From an operational point of view, so high a permeance ratio makes possible removal and insertion of pole pieces without danger of appreciable demagnetization. An iron shorting plug placed in gap 2 caused only a 3 percent drop in the magnetomotive force in gap 1, while the plug in gap 1 caused a 1 percent drop in gap 2.

An equally successful arrangement of magnets is produced where the direction of magnetization is perpendicular to the optical axis. Such an arrangement is shown schematically in Fig. 2, and photographically in Fig. 5. Two, four, or eight magnets may be utilized to give operating characteristics shown in Fig. 4 and

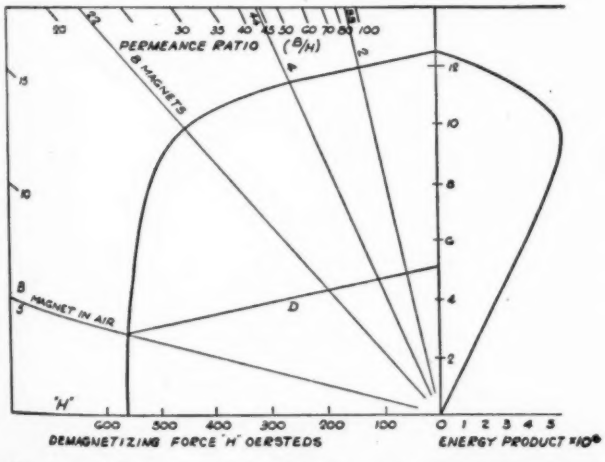


FIG. 4. B and H diagram for radially disposed magnets in groups of 2, 4, and 8. Demagnetization curve, Alnico V.

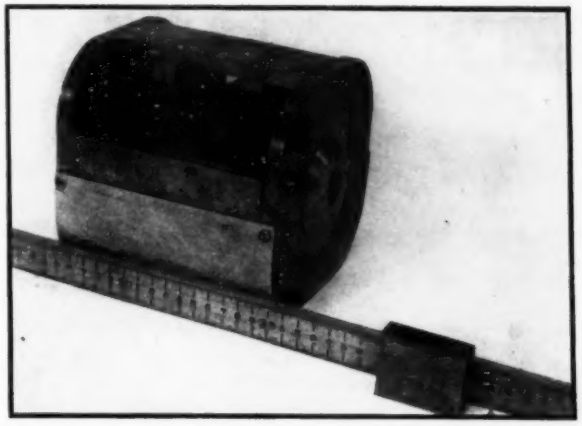


FIG. 5. A radial magnet system showing coils for energizing the magnets.

in Table I. Eight magnets are necessary to energize the pole pieces commonly used in electron microscopy. Actually, four magnets of approximately twice the cross-sectional area might be employed at the middle of the assembly. Magnets should always be employed in pairs so that radial symmetry may be observed. The magnets disposed perpendicularly to the axis do not introduce observable astigmatism in electron microscope images of about 200A resolution. Stray fields about the radial system are smaller than those about the coaxial system. Removal of a pole piece changes the permeance of the system on the order of one percent and hence represents but a small demagnetizing force.

The best mass utilization of magnet material is where the $B \times H$ curve is a maximum. However, for the application to electron microscopy, this is a minor consideration. Of more importance where pole pieces are to be changed is a circuital permeance ratio, large with respect to the value of the radio of the permeance of the pole piece to the permeance of the magnets. Such a condition prevents excessive demagnetization of the magnets when pole pieces are changed.

FABRICATION

Magnetization of magnetic material offers problems which should be solved during the design stage. It is necessary to magnetize magnet material while it is in position in its magnetic circuit in order to produce a magnet with the highest possible residual magnetism. Two means of magnetization are possible: (1) applying a MMF across the magnet; (2) passing a heavy dc current from the discharge of a capacitor bank through a few turns of wire wound around the magnet. The type

TABLE I.

Number of magnets	2	4	8
Gap strength (amp turns)	430	820	1500
Permeance ratio	88	44	22

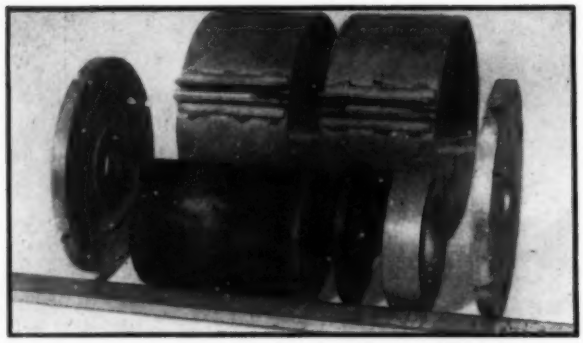


FIG. 6. Demountable test system for studying P.M. lens systems.

of assembly shown in Fig. 1 lends itself to magnetization with an external field applied to the spool ends. Alnico *V* requires 3000 oersteds for magnetization; hence, the magnets of Fig. 4 require 32,000 gilberts to be magnetized. However, there is an air gap of 1.9 centimeters in series with the magnet which must pass about 18,000 gauss. Thus, about 34,000 gilberts are necessary for the air gap drop. The total is about 60,000 gilberts, or 48,000 ampere turns. The iron shunt circuit due to the outer shell of the spool will have a magnetizing force of about 3000 oersteds, which will saturate it and permit the required MMF to appear across the magnet.

The assembly shown in Fig. 2 does not lend itself to magnetization by an external source. It is necessary to wind a coil of a few turns of No. 10 or No. 12 wire about each magnet of the assembly in the proper sense to produce the correct direction of magnetization. The coils are energized by flashing them with the discharge through an ignitron of a bank of capacitors, i.e., 2000 mf at 500 volts. All the coils are in series and are magnetized simultaneously. Once magnetized, magnets should never be removed from the magnetic circuit, because they will be partially demagnetized, as shown in Figs. 2 and 4, where the lines *B* represent the demagnetization curve for the long and short magnets in space. In Fig. 3 the line represents the condition of operation of the long axial magnets used in Fig. 1, when in the circuit. If the long magnet source is a single cylinder in space, it has a permeance ratio of 4. If magnetized externally and then placed into the assembly, it will operate at point *E* Fig. 3. If the long magnet is formed by four rectangular magnets, each magnet will have a permeance ratio of 14, and if magnetized externally and then operated in the magnet assembly, will operate at *D*. The operating condition for the magnets, when magnetized in position under similar external magnetization, is shown at *A*. Removal of magnets from the assembly, following magnetization in the assembly, causes them to operate under condition *B* or *C* of Fig. 3. Subsequent reinsertion will give operation at points *D* and *E* as described above. The line *D* of Fig. 4 shows the point of operation of externally energized radial magnets when used in groups of 2, 4, or 8 by its intersection with the corresponding permeance ratio line.

The fabrication of Alnico V magnets is sufficiently difficult to warrant the choice of simple shapes and small sizes. It must be remembered that Alnico V is not machinable and must be ground to final size. It has been found that the rectangular parallelopiped is the most suitable shape. Four such magnets placed edge to edge fulfill the function of the cylinder. Of more importance, they may be assembled after the lens spool has been fabricated and tested for vacuum tightness. Air gaps in series with the magnets should not exceed a few thousandths of an inch; hence, the pole ends should be ground parallel to within such tolerances.

It has been pointed out that with fixed magnetic strength, focusing of an electron beam may be accomplished by varying the electron accelerating voltage. It is also possible to obtain focusing by using an auxiliary coil wound about the spool core. Passing a small dc current through the coil will strengthen one gap and weaken the other, thus changing the effective focal lengths of the pole pieces in the gaps. The demagnetizing effect of the small supplementary coil has no appreciable effect upon the permanent magnet. Mechanical means for changing the reluctance of the circuits have been used to attain focus in other applications.[4] None of these means is as simple as varying the high voltage.

EXPERIMENTATION

The wide latitude inherent in the assumptions used in the design of magnetic circuits makes it wise to limit calculations to a rudimentary form and to determine conditions of operation with actual models. Since models suitable for use with vacuum systems, necessary for electron optical instruments, lack the flexibility desirable in experimentation, a completely

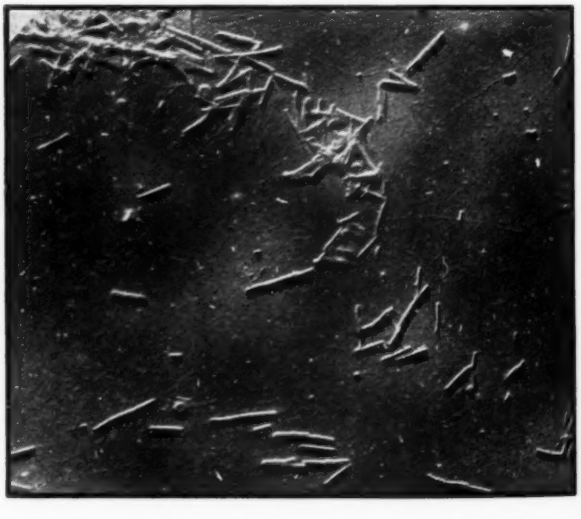


FIG. 7. Electron micrograph of tobacco mosaic virus showing individual particles. Diameter of the particles is 150A.

knock-down asembly, Fig. 6, was developed to provide the mechanical strength and magnetic circuits, while vacuum was maintained by operation of the entire assembly in a vacuum optical bench used for some time in this laboratory[5] for studying electron-optical systems. The holes in the plates were chosen to hold standard lens pole pieces whose properties were known. Studies with the above assembly led to the development of a vacuum tight lens useful on actual electron microscopes.

RESULTS

Each of the lenses shown in Figs. 1 and 2 has given high resolution pictures when used with electron microscope pole piece elements. Electronic magnifications estimated at well over ×10,000 have been obtained. Figure 7 is an actual micrograph showing a resolution of the order of 100A. There is no reason to believe that 100A represents the ultimate in resolution that is possible with the permanent magnet lens. As research is pressed on the problem, it is probable that full use will be made of the stability inherent in the permanent magnet lens energization, and the limit of resolution will be attained.

The stability of Alnico V is great. Lens assemblies such as those described above have been in use for two years without apparent loss of field strength. The uniformity of commercial magnets is such that several sensibly identical lenses have developed practically identical magnetomotive forces. Any radial asymmetry of the field has been below that which might cause an observable asymmetry in the microscope image of 100A resolution.

[5] J. Reisner and R. Picard, Rev. Sci. Instr. **19**, 556 (1948).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Internal Friction of Ferromagnetic Materials

A. FERRO AND G. MONTALENTI
Center for Electrophysics Studies of Galileo Ferraris National Electrical Institute, Torino, Italy

(Received October 10, 1950)

It is common knowledge that in cyclically stressed ferromagnetic materials an energy loss due to internal friction is observed; this part of loss disappears when the material is brought to saturation or above the Curie temperature. It is also known that a stress, applied to a ferromagnetic material, produces a motion of Weiss domains from their rest position.

The scope of this paper is to verify experimentally the following hypothesis: in cyclically stressed ferromagnetic materials, energy loss because of magneto-elastic internal friction is induced by a domain motion due to the applied stress itself.

The ratio I_r/I_s (where I_r=residual induction, and I_s=saturation intrinsic induction) has been chosen as an index of the domain position; its measure has been carried out by normal ballistic method on a permeameter rigged with a torsion applying device.

Energy loss was measured by means of a torsion pendulum; a coil allowed the saturation of specimen in order to separate magneto-elastic from purely mechanical losses. Plotting curves of I_r/I_s and of magneto-elastic losses as a function of applied stress and comparing them for each tested material, it has been found that they, besides having the same shape, show abrupt slope variations at practically the identical value of stress. It has further been observed that a specimen presenting only a low percent variation of I_r/I_s ratio (i.e., in which domains had slightly moved from their rest position) showed also moderate magneto-elastic losses. These results appear to confirm assumptions made about the magneto-elastic internal friction in ferromagnetic materials. The reason why motion of domains takes place in an essentially irreversible way still remains unknown.

INTRODUCTION

As it is well known, a metal does not behave as perfectly elastic; if stressed under mechanical tension, varying from a maximum positive value to an identical negative value, the stress-strain curve shows a closed cycle with positive area, which represents the energy loss per cycle of the specimen. Referred to the unit volume of the specimen, the energy loss measures, as stated, the internal friction of the material as a function of the maximum value of stress to which the specimen has been submitted. Materials which may be in both the nonmagnetic and ferromagnetic phase show in the latter a much higher internal friction than in the former. Becker and Kornetzki[1] observed that the internal friction of ferromagnetic materials decreases when the material is submitted to a magnetic field of intensity sufficient to saturate it. Zacharias[2] observed that internal friction in nickel decreases abruptly if the material is heated to a temperature above Curie's. These statements are confirmed by many other authors[3,4] who carried out studies about this matter. The domain theory[5] may be helpful to interprete the phenomena. As it is common knowledge, this theory foresees, and experience confirms, that a mechanical tension tends to turn the domains in a direction either parallel or normal to the tension itself, depending upon the sign of the magnetostriction (Fig. 1).

It does not appear erroneous to assume that high

[1] R. Becker and M. Kornetzki, Z. Physik **88**, 634 (1934).

[2] J. Zacharias, Phys. Rev. **44**, 116 (1933).
[3] C. Boulanger, Rev. met. **255**, (1949).
[4] C. Boulanger, Compt. rend. **226**, 1340 (1948).
[5] R. M. Bozorth, Z. Physik **124**, 519 (1948).

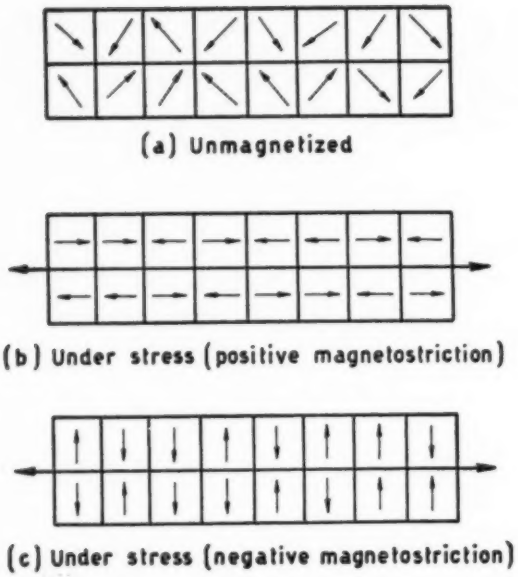


FIG. 1. Effect of stress on domain orientation.

value of internal friction in ferromagnetic materials arises from the domain motion starting from their rest position. This motion must be assumed as being associated with an energy loss. The above facts, observed by Becker, Kornetzki, and Zacharias, are qualitatively explained by this hypothesis.

The scope of the present paper is to submit to a more direct experimental inspection the above mentioned hypothesis by comparing the internal friction values in terms of mechanical stress, with the ratio I_r/I_s = residual induction to saturation intrinsic induction, which is a suitable index of the domain motion, starting from their rest position. In order to realize how this ratio suits the purpose, we will consider a sphere having arbitrary radius and vectors parallel to the domains oriented in that direction. Should a preferential direction not exist, the number of domain per solid angle unit is constant. Figure (2a) shows the position of the domains of anisotropic material in three different magnetization states: (a) not magnetized, (b) saturated, and (c) with residual magnetization. It is easy to show that, if we indicate with I_r the residual magnetization intensity and with I_s the intrinsic magnetization intensity, the I_r/I_s ratio, in this particular case, is equal to 0.5.

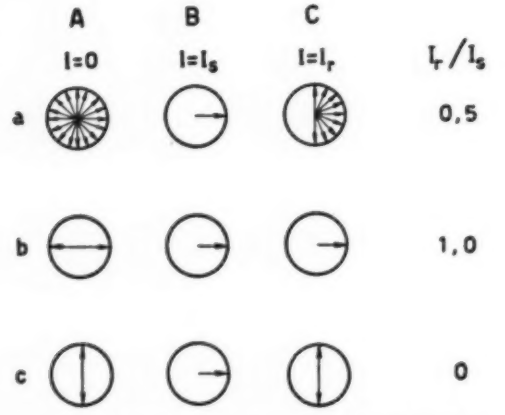


FIG. 2. Values of ratio I_r/I_s for three different domain distributions.

If the material is stressed to a sufficiently high stress, I_s keeps its constant value while, depending upon the respective sign of the stress (tension or compression) and of the magnetostriction, the domains will tend to settle, following the diagram (*b*) or (*c*) of Fig. (2), with a corresponding variation in residual induction.[6,7] This simple case shows how the ratio I_r/I_s can be assumed as being an index of the domain movements from rest position.

The internal friction, measured under a given stress amplitude, is the sum of internal frictions due to several causes: some depending on oscillation frequency and mainly due to thermal diffusion, components diffusion, and viscous relaxation; and some others, not depending on frequency, due to irreversible phenomena of interatomic nature introduced by the stress. In order to consider for the internal friction only the part due to the magnetic-elastic losses, we will consider only the difference between the total value of internal friction and the part measured when the material is magnetically saturated. This difference is referred to as "magneto-elastic internal friction." As oscillation frequencies to which the specimens have been submitted are not higher than a few cycles per second, the above losses cannot have been caused by Foucault currents having identical frequency.

Let us now consider the curve obtained by plotting the magneto-elastic internal friction as a function of maximum stress and compare it with the analogous curves referred to the values of I_r/I_s ratio also as a function of stress. If both curves show a similar form, and if the points where the slope changes abruptly are found to correspond for a certain value of the stress, it may be said that the hypothesis offered on the origin of internal friction for ferromagnetic materials is correct.

MEASURE METHODS

If free oscillations are impressed to a torsion pendulum consisting of a hollow or solid cylindrical specimen clamped at one end and having a flywheel at the other, it obviously results that the higher the internal friction the higher the decrease in amplitude of oscillation.

The mean energy loss value in one cycle per unit volume in the specimen is then

$$E = K\delta\tau^2/G,$$

where δ = logarithmic decrement of oscillation, τ = tangential stress at the edge of the test piece, G = tangential elasticity modulus, and K is a constant depending on stress distribution in the specimen: in our case, for hollow specimen $K=0.625$ and for solid one $K=0.5$.

[6] M. Kornetzki, Z. Physik 560 (1943).
[7] Jl. Snoek, Physica 8, 745 (1941).

TABLE I.

Specimen No.	Dimensions of the test piece mm: Ext. diam	Int. diam	Length	Material	Composition	Heat treatment	Saturation intrinsic induction Wb/m²	Coercive force Asp/m	I_r/I_s initial value
1	6	3	45	Nickel	Ni 99.5%	As cast	0.61	700	0.52
2	8	0	65	Steel	C=0.38, Ni=2.2, Cr=0.6	Quenched in oil and tempered for 1 hr at 650°C	1.91	1.200	0.74
3	8	0	65	Steel	id.	id.	1.92	1.200	0.73
4	6	3	45	Steel	id.	Quenched in oil tempered for 1 hr at 520°C	1.91	1.300	0.77
5	6	3	45	Steel	C=0.4	Quenched in water tempered for 3 hr at 650°C	1.97	1.000	0.67
6	6	3	45	Steel permanently strained in torsion 4 percent	C=0.38, N=2.2, Cr=0.6	The same annealed after work hardening at 500°C for 10 min	1.83	1.600	0.63

The experimental equipment containing the torsion pendulum was similar to that already used by Föppl,[8] Locati,[9] Boulanger,[3] and others.

Impressing to the specimen a certain initial torsion and releasing the flywheel causes the system to perform damped oscillations, which are recorded on photographic paper by means of an optical device.

The measurements can be repeated with satisfactory approximation, and the average result scattering is about ±5 percent.

For each specimen, the total energy loss, i.e., the loss due both to the magneto-elastic part and to the viscous relaxation and termo-elastic effect, has been determined. After this first determination, the material has been brought to saturation by means of a strong magnetic field (60000 Asp/m) and the measurements of damping repeated. In such conditions there is no magneto-elastic loss and, therefore, loss due to viscous friction only can be measured. The latter, substracted from the total energy loss previously found, shows directly the magneto-elastic loss.

The test piece consists of a cylindrical bar, clamped at both ends, one to a large free flywheel and the other to a heavy bench; the dimensions and the characteristics of the various specimens are shown in Table I.

Measurements of I_s and I_r have been carried out by means of a normal ballistic method, having placed the specimen in a Neumann's permeameter particularly suitable for measurements in strong magnetic fields.[10]

It may be assumed that the error in the I_r/I_s ratio is of the order of ±2 percent. To obtain the required value of shearing stress in the specimen, one end is rigidly clamped, and a torque is applied to the other end by means of a weight at the end of a lever.

The static stress, to which the specimen has been submitted, being torsional, the I_r/I_s ratio and the magneto-elastic internal friction curves are, under this point of view, exactly comparable to each other; besides, it is to be observed that both values represent the mean value for the whole test piece.

EXPERIMENTAL RESULTS

Tests have been carried out on cast nickel (Spec. $n°$ 1), on quenched and tempered steel ($n°$ 2), ($n°$ 3), ($n°$ 4), and ($n°$ 5); and finally, on a permanently strained specimen of the same steel ($n°$ 6).

The curves ①, ②, and ③ of each diagram (Fig. 3) represent, respectively, energy losses due to magneto-elastic friction (full line), the nonmagnetic or viscous part of internal friction (dotted line), and the I_r/I_s ratio (dotted dashed line), all in terms of applied stress.

Generally, for each specimen, a good qualitative correspondence is found between I_r/I_s and the energy loss curves. In fact, both curves show a first portion nearly parallel to the abscissa axis; a second one, the inclination of which increases considerably; and a last one, which again is nearly parallel to the abscissa axis. In particular, we notice that the abscissa of the point, where magneto-elastic losses become noticeable, is very near to the abscissa of the point where the first variation of the I_r/I_s is noticed.

It must be kept in mind, in comparing the curves, that, from energy considerations, the relation holding between losses and I_r/I_s ratio variation is an even function and, as a first approximation, a second degree function; this can be easily verified from the data referred.

In the same way, in those cases where measurements

[8] O. Föppl, Iron and Steel Inst. Meeting (1936).
[9] L. Locati, Met. ital. 32, 235 (1940).
[10] H. Neumann, A.T.M. J66-2 (November, 1943).

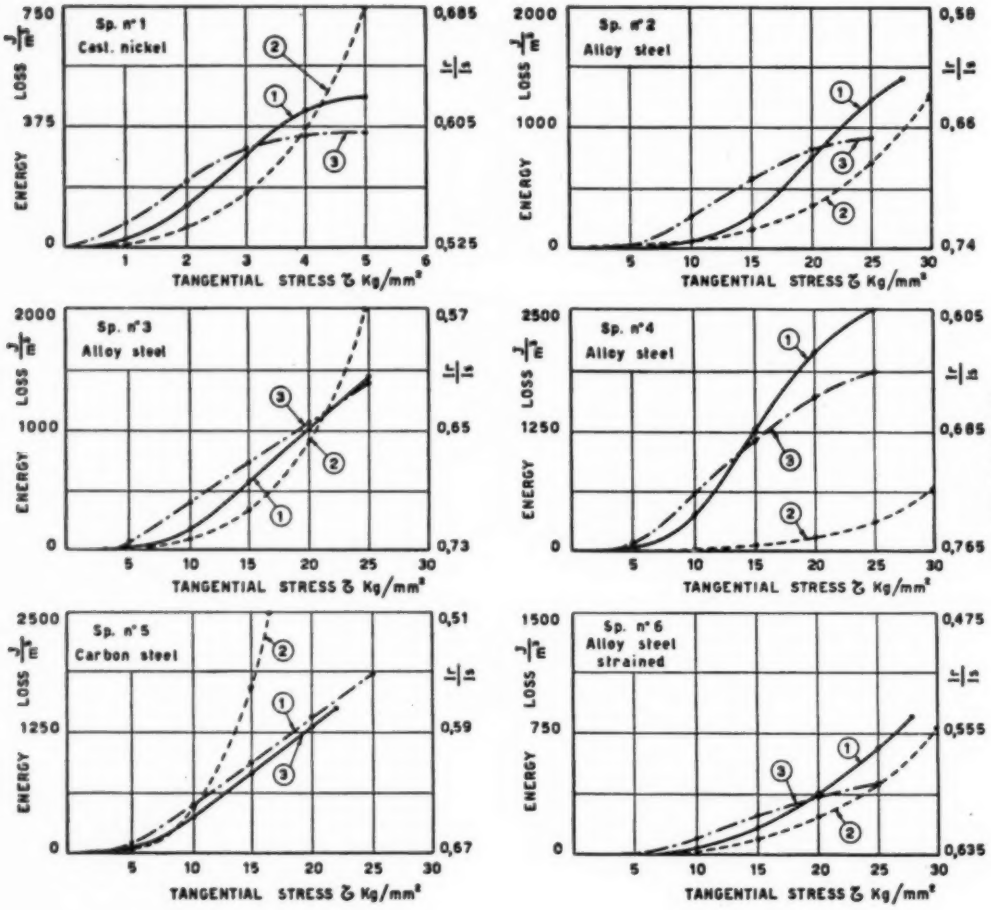


FIG. 3. Relation of magneto-elastic hysteresis to stress induced changes in magnetic remanence $-\cdot-\cdot-\cdot-I_r/I_s$. ——— Magneto-elastic energy losses. - - - - - - - - - - Non-magnetic energy losses.

have been possible, the elastic limit being sufficiently high, we may observe that the point where the magneto-elastic losses become constant almost coincides with the point where the I_r/I_s ratio also becomes constant.

At last the result obtained from the strained specimen is of particular interest. The test piece has been permanently twisted to a deformation equivalent to 4 percent plastic strain and has been annealed for 10 minutes at 500°C. In this case, it has been observed that magneto-elastic losses began at a much higher stress and, at the same stress, they were equal to about one-fourth of those of the same specimen before straining. At the same time, I_r/I_s variation, measured under torsional stress in both directions, remains very small. It thus appears that, when domains move only slightly from the rest position, a decrease in the magneto-elastic internal friction is noticed.

CONCLUSION

The variation of I_r/I_s ratio in terms of applied stress is a fair index of the variations in the direction of Weiss domains. The shape of its curve is decidedly similar to the one of magneto-elastic internal friction loss for all materials. In one permanently strained specimen,[6] where it was noticed that the I_r/I_s ratio varied only a small percent with the stress, the correspondent value of internal friction is very low. The results of these different experiments then allow for the conclusion that internal friction of ferromagnetic materials is due to the motion of Weiss' domains, which must take place with energy loss if experimental results are to be explained; the question about the study of the causes, owing to which magneto-elastic motion of the domains takes place with loss of energy, is still open.

ACKNOWLEDGMENTS

We thank Professor Zerbini, Chief of the Magnetic Section of the National Electrical Institute, for his precious advice during the preparation and accomplishment of the present work.

The Electric Tunnel Effect across Thin Insulator Films in Contacts

RAGNAR HOLM
Stackpole Carbon Company, St. Marys, Pennsylvania
(Received August 21, 1950)

Existing calculations on the tunnel effect through thin films in contacts apply to very weak and very strong electric fields. It is the main purpose of the present paper to complete the picture by the treatment of intermediate cases, which are important for many contact applications.

INTRODUCTION

THE electric tunnel effect through films plays a considerable role in the behavior of contacts, and several papers have been published on this subject. In these publications, the calculations have been simplified by limiting them either to very weak or to very strong fields, but, unfortunately, due attention to these limitations is missing in several applications. It is the main purpose of the present paper to complete the picture by the treatment of intermediate cases. A mathematical approach has been chosen which is similar to that of Sommerfeld and Bethe.[1-3] These authors use cgs units, and in order to facilitate reference to their paper, we also perform our deductions in cgs units; yet, since practical problems are expressed better in amperes, volts, and A units, some important equations written in these units are listed in Sec. V. They are graphically illustrated in Figs. (6) and (7).

For extreme (weak or strong) fields, the solutions have been completed to include the influence of the image force.[1-3] Though we simplify our calculations for intermediate fields by neglecting the image force, the result will have a general application, since it shows how the intermediate states interpolate between the extremes. We shall indicate an approximate method of working the interpolation when the image force is taken into account.

SYMBOLS

x, distance, cm or A.
s, thickness of film.
m, mass of electron, $m=9.1\times10^{-28}$g.
$v=(v_x^2+v_y^2+v_z^2)^{\frac{1}{2}}$ velocity of electrons in cm/sec.
ϵ, charge of electron, 1.6×10^{-19} coulomb.
h, Planck's constant, $h=6.625\times10^{-27}$ erg×sec $=4.14\times10^{-15}$ electron volt×sec.
k, Boltzmann's constant, $k=1.38\times10^{-16}$ erg×deg^{-1} $=8.69\times10^{-5}$ electron volt×deg^{-1}.
T, temperature in °K.
j, current density in amp/cm^2.
σ, tunnel resistivity in ohm cm^2 (see Sec. II).
V, voltage across film.
F, strength of electric field, in v/cm or v/A.
E, energy level of electrons.
$E_x=\frac{1}{2}mv_x^2$.
$D(E_x)$, transmission coefficient for electrons which are characterized by E_x.
φ, work function in ergs or electron volts.
η, Fermi level.
γ, bottom of conducting band of a semiconductor.
ϕ, function measured in the same unit as φ or η.

I. SOME FUNDAMENTAL EQUATIONS

Only symmetric contacts shall be considered with equal metallic or semiconducting members. The contact interface is supposed to contain a film of uniform thickness s (in the x direction), and the action of the film on electrons will be represented by a potential hill, the height of which at a point x may be $\phi(x)$ ergs over the zero level of the metal. To the velocity component v_x of the electron, we correlate an energy component $E_x=\frac{1}{2}mv_x^2$. Then, the probability for this electron to penetrate the hill is given by the transmission coefficient $D(E_x)$ for which we shall use the well-known approximate expression[4,5]

$$D(E_x)=\exp\left[-(4\pi/h)\int_0^s\{2m(\phi(x)-E_x)\}^{\frac{1}{2}}dx\right]. \quad (1)$$

The tunnel effect may transfer N_1 electrons/cm^2 sec from the one, say the left contact member, and N_2 from the right one, making the observable electron current $N=N_1-N_2$. Denoting the electron distribution on elements dv_x, dv_y, dv_z by

$$f(E)dv_xdv_ydv_z, \quad \text{where} \quad E=\tfrac{1}{2}mv^2,$$

we have

$$N_1=\int_0^{E_m}v_xD(E_x)dv_x\int_{-\infty}^{\infty}\int \times f(E)dv_ydv_z \text{ electrons/cm}^2\text{ sec};$$

E_m is defined below. A corresponding expression holds for N_2. Calling the voltage across the film V and con-

[1] A. Sommerfeld and H. Bethe, *Handbuch der Physik von Geiger und Scheel* (Verlag. Julius Springer, Berlin, 1933), 2. Aufl. Vol. 24/2, pp. 450f.
[2] R. Holm, *Electric Contacts* (H. Geber, Stockholm, 1946), §22 and 23.
[3] L. Nordheim, Physik. Z. 30, 177 (1929).
[4] See reference 1, paragraphs 19–21.
[5] V. Rojansky, *Introductory Quantum Mechanics* (Prentice-Hall, Inc., New York, 1938), §34, and 38.

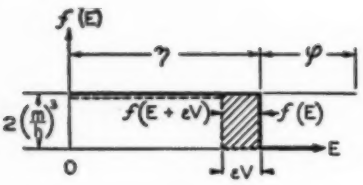


FIG. 1. Distribution functions $f(E)$, solid lines, and $f(E+V)$, dashed lines, at $T=0$, plotted against energy levels E.

sidering that $D(E_x)$ is the same in both directions, we obtain

$$N=\int_0^{E_m} v_x D(E_x)dv_x \int_{-\infty}^{\infty}\int \times[f(E)-f(E+\epsilon V)]dv_y dv_z \text{ electrons/cm}^2\text{ sec.} \quad (2)$$

Here we have neglected the distortion of $f(E)$ by the current.[6] The upper limit for the first integral should lie below the surface potential of the left contact member, since electrons above this level are free to overcome the potential hill without the aid of the tunnel effect. However, the number of electrons at higher levels is so small, that we are largely free in choosing this upper limit. Provisionally, we denote the value chosen for this limit by E_m.

We introduce

$$\rho^2=v_y^2+v_z^2, \quad \tan\theta=v_z/v_y, \quad \text{and} \quad E_{yz}=\tfrac{1}{2}m\rho^2.$$

Then

$$N=(2\pi/m^2)\int_0^{E_m} D(E_x)dE_x \int_0^{\infty} \times[f(E)-f(E+\epsilon V)]dE_{yz}$$

or, since integrating along the path of dE_{yz} from $E_{yz}=0$ means the same as integrating along the path of dE from $E=E_x$,

$$\left.\begin{aligned} N&=(2\pi/m^2)\int_0^{E_m} D(E_x)Y(E_x)dE_x \\ \text{where}\quad Y(E_x)&=\int_{E_x}^{\infty}[f(E)-f(E+\epsilon V)]dE \end{aligned}\right\}. \quad (3)$$

Simplifying in the same manner as is done in reference 1, we use f values which belong to zero absolute temperature and, for *metals*, obtain a distribution picture;

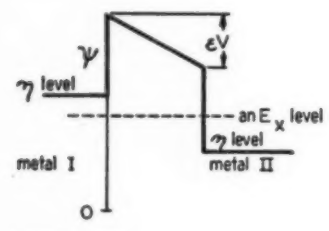


FIG. 2. Potential hill representing a contact film between contact members I and II of the same metal when the potential V is applied between them.

see Fig. 1, with the notations, η = Fermi level, φ = work function for electron emission from the metal to the film, V = potential drop across the film.

The general shape of the potential hill is illustrated in Fig. 2. The case of a film between semiconducting contact members is treated in Sec. V.

II. SMALL V-VALUES

For very "small V values," Sommerfeld and Bethe put

$$\begin{aligned} Y(E_x)&=2(m/h)^3\epsilon V \quad \text{for} \quad E_x\leqq\eta-(\epsilon V/2), \\ Y(E_x)&=0 \quad\quad\quad\quad \text{for} \quad E_x>\eta-(\epsilon V/2), \end{aligned} \quad (4)$$

and

$$E_m=\eta-(\epsilon V/2).$$

They deduce

$$D(E_x)\approx\exp[-(4\pi s/h)\{2m(\eta+\varphi-(\epsilon V/2)-E_x)\}^{\frac{1}{2}}], \quad (5)$$

and, after some manipulations and simplifications, arrive[7] at Eq. (6).

$$N=\frac{\epsilon V}{4\pi hs^2}\Big[(A\varphi^{\frac{1}{2}}+1)\exp(-A\varphi^{\frac{1}{2}}) -\{A(\eta+\varphi-\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}+1\}\exp[-A(\eta+\varphi-\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}] \quad (6)$$

with

$$A=(4\pi s/h)(2m)^{\frac{1}{2}}.$$

The second term within the bracket [ ] is usually negligible, and $1\ll A\varphi^{\frac{1}{2}}$.

In the following we are not interested directly in the number N but in the ratio σ between the voltage V and the density of the tunnel current. σ has the dimension of a resistance per cm² and may be called the *tunnel resistivity* of the film. We have

$$\sigma=V/\epsilon N, \quad (7)$$

and, according to Eq. (6), taking advantage of the indicated approximations,

$$\sigma\approx(h^2s/\epsilon^2(2m\varphi)^{\frac{1}{2}})\exp[(4\pi s/h)(2m\varphi)^{\frac{1}{2}}]. \quad (8)$$

III. EXTREMELY LARGE V VALUES

"Extremely large V values" here means

$$\epsilon V\geqq\eta+\varphi.$$

Figure 3 illustrates the potential hill for this case. The current from metal II is negligible,

$$Y(E_x)=\int_{E_x}^{\infty} f(E)dE,$$

and, according to Eq. (3),

$$N=(2\pi/m^2)\int_0^{\eta} D(E_x)dE_x \int_{E_x}^{\infty} f(E)dE$$

[6] See reference 1, the discussion of J_1 and J_2 on p. 449.

[7] See reference 1, Eq. (21.12).

with

$$f(E)=2(m/h)^3 \quad \text{for} \quad E\leqq\eta,$$
$$f(E)=0 \quad \text{for} \quad E>\eta;$$

hence,

$$\int_{E_x}^{\infty} f(E)dE=2(m/h)^3(\eta-E_x).$$

As for $D(E_x)$, a simple integration gives

$$D(E_x)\approx\exp[-(8\pi(2m)^{\frac{1}{2}}x/3h)(\eta+\varphi-E_x)^{\frac{1}{2}}],$$

or, since

$$x/(\eta+\varphi-E_x)=s/\epsilon V,$$
$$D(E_x)\approx\exp[-(8\pi(2m)^{\frac{1}{2}}/3h)(s/\epsilon V)(\eta+\varphi-E_x)^{\frac{1}{2}}]. \quad (9)$$

We approximate $(\eta+\varphi-E_x)^{\frac{1}{2}}$ by $\varphi^{\frac{1}{2}}[1+\frac{3}{2}(\eta-E_x)/\varphi]$ and take into account that the field strength F satisfies

$$F=V/s \quad (10)$$

which is the same for any E_x. We finally obtain

$$N=\frac{4\pi m}{h^3}\exp\left[-\frac{8\pi}{3hsF}(2m)^{\frac{1}{2}}\varphi^{\frac{3}{2}}\right]\int_{\eta-E_x=0}^{\eta}(\eta-E_x)$$
$$\times\exp\left[-\frac{4\pi}{h\epsilon F}(2m\varphi)^{\frac{1}{2}}(\eta-E_x)\right]d(\eta-E_x). \quad (11)$$

Following the procedure of Sommerfeld and Bethe, we approximate by integrating to ∞ instead of to η and find

$$N=\frac{\epsilon^2F^2}{8\pi h\varphi}\exp\left[-\frac{8\pi}{3h\epsilon F}(2m)^{\frac{1}{2}}\varphi^{\frac{3}{2}}\right]\text{electrons/cm}^2\,\text{sec} \quad (12)$$

which, except for Sommerfeld's and Bethe's consideration of the image force, corresponds to their equation (19.24) (see reference 1).[3]

IV. INTERMEDIATE V VALUES

"Intermediate V values" means $0<\epsilon V<\eta+\varphi$. We start from Eq. (3) and distinguish two regions for the integration of this equation; see Fig. 4,

$$\text{I with } E_x<\eta+\varphi-\epsilon V$$
$$\text{II with } E_x\geqq\eta+\varphi-\epsilon V. \quad (13)$$

Region II is concerned only if $\epsilon V>\varphi$. For region I, D may be expressed according to Eq. (5). The expression for the number of electrons N_I becomes similar to Eq. (6), but with the following difference. We approximate by putting

$$Y(E_x)=2(m/h)^3\epsilon V \quad \text{for} \quad E_x<\eta-\epsilon V,$$
$$Y(E_x)=0 \quad \text{for} \quad E_x>\eta, \quad (14)$$
$$Y(E_x)=2(m/h)^3(\eta-E_x) \quad \text{for} \quad \eta-\epsilon V<E_x<\eta$$

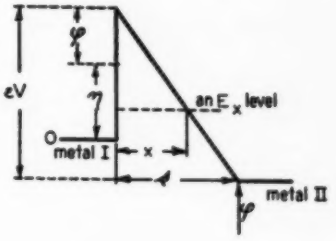


FIG. 3. The potential hill representing the film when $\epsilon V>\eta+\varphi$.

and obtain

$$N_I=(4\pi m/h^3)\epsilon V\int_0^{\eta-\epsilon V}$$
$$\times\exp[-A(\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x)^{\frac{1}{2}}]dE_x$$
$$+(4\pi m/h^3)\int_{E_x=\eta-\epsilon V}^{\eta}(\eta-E_x)$$
$$\times\exp[-A(\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x)^{\frac{1}{2}}]dE_x, \quad (15)$$

where

$$A=(4\pi s/h)(2m)^{\frac{1}{2}}.$$

The first term of the expression for N_I is nearly similar to the right member of Eq. (6). The second term may be written

$$(4\pi m/h^3)\int_{E_x=\eta-\epsilon V}^{\eta}(\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x)$$
$$\times\exp[-A(\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x)^{\frac{1}{2}}]dE_x$$
$$-(4\pi m/h^3)(\varphi-\tfrac{1}{2}\epsilon V)\int_{E_x=\eta-\epsilon V}^{\eta}$$
$$\times\exp[-A(\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x)^{\frac{1}{2}}]dE_x.$$

For the evaluation of the first integral of this second term, we write

$$z^2=\eta+\varphi-\tfrac{1}{2}\epsilon V-E_x$$

and apply

$$\int z^3e^{-Az}dz=-e^{-Az}\left[\frac{z^3}{A}+\frac{3z^2}{A^2}+\frac{6z}{A^3}+\frac{6}{A^4}\right].$$

With reasonable approximations, we finally find that

$$N_I=\frac{(2m)^{\frac{1}{2}}}{h^2s}\epsilon V(\varphi+\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}\exp[-A(\varphi+\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}]$$
$$+\frac{(2m)^{\frac{1}{2}}}{h^2s}\frac{3}{A}\{(\varphi-\tfrac{1}{2}\epsilon V)\exp[-A(\varphi-\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}]$$
$$-(\varphi+\tfrac{1}{2}\epsilon V)\exp[-A(\varphi+\tfrac{1}{2}\epsilon V)^{\frac{1}{2}}]\}. \quad (16)$$

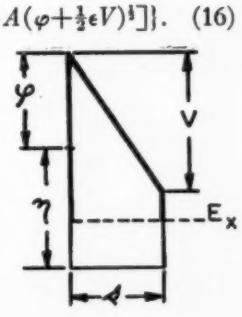


FIG. 4. The potential hill representing the film when $\epsilon V<\eta+\varphi$.

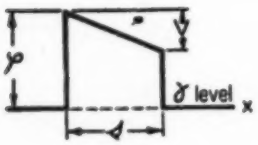


FIG. 5. Initially rectangular potential hill deformed by the application of the potential V between the contact members.

For region II, where $\epsilon V > \varphi$, the calculation will be similar to that in Sec. III, only with the difference that the upper limit of integration in Eq. (11) must be changed to $\epsilon V - \varphi$ instead of ∞. This adds a factor $(1-X)$ to the expression for N_{II}, where X is given below in Eq. (28). Hence,

$$N_{II} = \frac{\epsilon^2 F^2}{8\pi h\varphi}(1-X) \times \exp\left[-\frac{8\pi}{3h\epsilon F}(2m)^{\frac{1}{2}}\varphi^{\frac{3}{2}}\right] \text{electrons/cm}^2 \text{ sec.} \quad (17)$$

The total electron current from metal I to metal II is

$$N = N_I + N_{II}. \quad (18)$$

V. THE TUNNEL EFFECT IN AN INSULATING CONTACT FILM WHEN BOTH CONTACT MEMBERS CONSIST OF THE SAME SEMICONDUCTING MATERIAL

We confine the calculations to a work function φ and temperatures which satisfy $\varphi \gg kT$ and, moreover, to the case of exclusively extrinsic conduction. Considerable simplifications can then be justified. The occupied levels in the conducting band will differ from the lower boundary γ of this band by amounts of the order kT; and we may consider the approximate assumption that all n conducting electrons (per cm^3) in the member with the higher (negative) potential belong to the γ level. We neglect the image force and represent the potential hill by Fig. 5. A simple integration yields

$$D(E_x) = \exp[-(8\pi(2m)^{\frac{1}{2}}/3h)(s/\varphi) \times \{\varphi^{\frac{3}{2}} - (\varphi-\epsilon V)^{\frac{3}{2}}\}]. \quad (19)$$

The number of electrons impinging on the film from the left member is, according to the kinetic theory,

$$N_l = n(kT/2\pi m)^{\frac{1}{2}} \text{ electrons/cm}^2 \text{ sec.} \quad (20)$$

The number impinging from the right member is

$$N_r = N_l e^{-\epsilon V/kT}.$$

Consequently [see Eq. (3)],

$$Y(\gamma) = N_l - N_r = n(kT/2\pi m)^{\frac{1}{2}}(1-e^{-\epsilon V/kT}).$$

Inserting this result in Eq. (3) and observing that the integration here is the same as multiplying by 1, we obtain

$$N = n\left(\frac{kT}{2\pi m}\right)^{\frac{1}{2}}(1-e^{-\epsilon V/kT}) \times \exp\left[-\frac{8\pi(2m)^{\frac{1}{2}}}{3h}\frac{s}{\epsilon V}\{\varphi^{\frac{3}{2}} - (\varphi-\epsilon V)^{\frac{3}{2}}\}\right]. \quad (21)$$

Here, for $V < 0.5$ volt, the exponential function may be simplified to

$$\exp[-(4\pi s/h)\{2m(\varphi - \tfrac{1}{2}\epsilon V)\}^{\frac{1}{2}}]. \quad (22)$$

The error thus introduced is less than 1 percent.

Equations (21) and (22) are known. Equation (21) or very similar equations were used, for instance, by Braun and Busch,[8] by Miss Dilworth,[9] and, with reference to the latter, by Mitchell and Sillars.[10] The reason for repeating the short deduction is to point out simplifications involved.

VI. FILM RESISTIVITY EXPRESSED IN PRACTICAL UNITS

For numerical calculations, it is more convenient to express the current in amperes, V, η, and φ all in volts (ϵ being included in the numerical constants), s in A units, and F in volt/A. The following formulas for the current density j are obtained by converting Eqs. (6), (12), (16), and (17) to these units. We find: for equal *metallic* contact members, using the Sommerfeld-Bethe approximation

for $V \to 0$,

$$\sigma = 3.18\times10^{-11}(s/\varphi^{\frac{1}{2}}) \exp(1.025 s\varphi^{\frac{1}{2}}) \text{ ohm cm}^2; \quad (23)$$

for $V \gtrless \eta + \varphi$,

$$\epsilon N = j = 1.53\times10^{10}(F^2/\varphi) \times \exp(-0.68\varphi^{1.5}/F) \text{ amp/cm}^2; \quad (24)$$

and

$$\sigma = (V/j) = (Fs/j) = 6.54\times10^{-11}(\varphi/F)s \times \exp(0.68\varphi^{1.5}/F) \text{ ohm cm}^2. \quad (25)$$

With $V < \eta + \varphi$,

$$j_I \approx (10^{11}/3.18s)V(\varphi+\tfrac{1}{2}V)^{\frac{1}{2}} \exp[-1.025s(\varphi+\tfrac{1}{2}V)^{\frac{1}{2}}] + (10^{11}/1.06s^2)\{(\varphi-\tfrac{1}{2}V)\exp[-1.025s(\varphi-\tfrac{1}{2}V)^{\frac{1}{2}}] - (\varphi+\tfrac{1}{2}V)\exp[-1.025s(\varphi+\tfrac{1}{2}V)^{\frac{1}{2}}\} \text{ amp/cm}^2, \quad (26)$$

$$j_{II} = 1.53\times10^{10}(F^2/\varphi) \times \exp(-0.68\varphi^{1.5}/F)(1-X) \text{ amp/cm}^2, \quad (27)$$

where

$$X = [1.025 s\varphi^{\frac{1}{2}}(1-\varphi/V)+1] \times \exp[-1.025 s\varphi^{\frac{1}{2}}(1-\varphi/V)]; \quad (28)$$

and, according to Eqs. (13),

$$V > \varphi.$$

X is usually a small quantity which may be neglected. Equation (27) then becomes equal to Eq. (24). Finally,

$$\sigma = V/(j_I + j_{II}) \text{ ohm cm}^2. \quad (29)$$

---

[8] A. Braun and G. Busch, Helv. Phys. Acta **15**, 571 (1942), particularly Eq. (1).

[9] C. C. Dilworth, Proc. Phys. Soc. (London) **60**, 315 (1948), particularly Eq. (9).

[10] E. W. J. Mitchell and R. W. Sillars, Proc. Phys. Soc. (London) **62B**, 509 (1949).

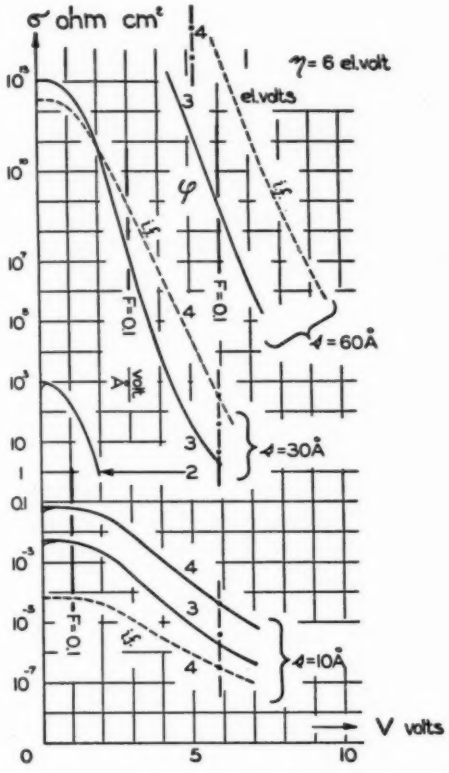


FIG. 6. Tunnel resistivity σ plotted against the voltage V across the film in a metallic contact. Values of η, φ, and s are indicated on the figure.

For equal, semiconducting contact members

$$\sigma=\frac{10^{14}\exp[0.683(s/V)\{\varphi^{\frac{1}{2}}-(\varphi-V)^{\frac{1}{2}}\}]}{2.5nT^{\frac{1}{2}}(1-e^{-11600V/T})}, \qquad (30)$$

where 0 is to be substituted for $(\varphi-V)^{\frac{1}{2}}$ when $V\geqq\varphi$. With $V<0.5$, the exp function may be simplified* to

$$\exp 1.025s(\varphi-\tfrac{1}{2}V)^{\frac{1}{2}}. \qquad (31)$$

The term $e^{-11600V/T}$ is usually negligible.

VII. THE INFLUENCE OF THE IMAGE FORCE

The image force causes a rounding off of the edges of the potential hill, and thus diminishes its average height and increases the tunnel current. Sommerfeld and Bethe[1] have shown how to correct the formula (8) to take the image force into account. Their correction is based on approximating the hill by a symmetric parabola. Holm and Kirschstein[11] use the same method, with the difference that they fit the parabola closer to the assumed potential hill. We refer to their result (plainly reported and illustrated in reference 2 §22), in the following. The main correction for the image force consists in subtracting $(\epsilon^2/s)\ln 2$ ergs $=10/s$ electron volts from the work function φ in the exponent.[12] However, the coefficient in the exponent is also changed. The following example indicates the method which has been used for the calculation of those curves in Fig. 6, which have been corrected for the image force.

We want to find σ in the case of $\varphi=4$ electron volts, $s=30$A, $V=3$ volts, taking into account the image force. First, we calculate the exponential function in Eq. (22.03) of reference 2 and find, using the symbols of reference 2,

$$e^{A\phi}, \text{ with } A\phi=0.87s(\varphi-10/s)^{\frac{1}{2}}.$$

This indicates that we should substitute 0.87 instead of 1.025 in the exponents of Eq. (26). Another change is the replacing of φ by $\varphi-10/30$. The result of the calculation is

$$\sigma=3.8\times10^{8} \text{ ohm cm}^2.$$

VIII. NUMERICAL EVALUATIONS

Figure 6 illustrates the tunnel resistivity in a contact with equal *metallic members*. The values of φ in electron volts and of s in A units are indicated in the figure. The calculations have been made according to Eqs. (26), (27), and (29).

The dashed curves are marked "*i.f.*" with reference to "image force." They represent the consideration of the image force and have been calculated for extreme V values, according to Eqs. (22.03) and (23.04) of reference 2, and for the intermediate cases, according to the rule given in the last part of Sec. VI, with $\eta=6$ electron volts.

An obvious feature of all these curves is their horizontal asymtote for $V\to0$. This is probably a designation of the real physical phenomenon, but our equations actually define a little slope down to the σ-axis

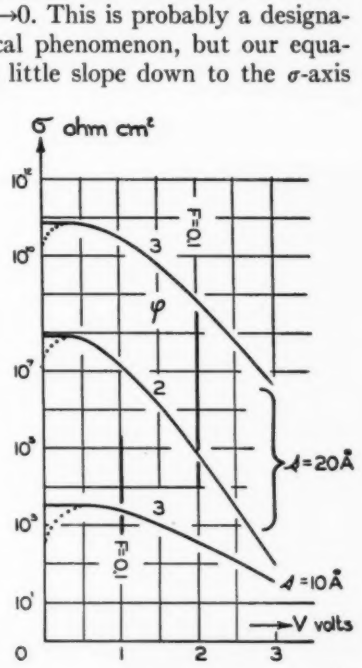


FIG. 7. Tunnel resistivity σ plotted against the voltage V across the film in a semiconducting contact. Values of φ and s are indicated on the figure.

* See remark at Eq. (22).

[11] R. Holm and B. Kirschstein, Z. tech. Physik **16**, 488 (1935).

[12] R. Holm, *Electric Contracts* (H. Geber, Stockholm, 1946), Eq. (22.03).

from a small maximum at about $V=0.25$ volt. This inaccuracy of the equations seems to result from the approximate character of the formula (1) for $D(E_x)$.

Holm and Meissner[13] measured σ in several contacts with very thin films. For instance, for $s\approx 5$A they did not find any clear deviation from the ohmic σ with V up to 0.035 volt.

Another inaccuracy of the formulas is manifested by the following fact. Let σ_{26} be a value of σ which is obtained from Eq. (26) for $V\to 0$, and let σ_8 be defined by Eq. (8). We would expect $\sigma_{26}=\sigma_8$ but find

$$\frac{\sigma_{26}}{\sigma_8}=\frac{1}{2.5-(s/4)(V/\varphi^{\frac{1}{2}})-(3.2/s\varphi^{\frac{1}{2}})}. \tag{32}$$

The deviation of the right member from 1 mainly indicates that the choice of E_m in Eqs. (4) is too rough, but at the same time it is typical for the inaccuracy of the fundamental formulas.

Figure 7 illustrates the tunnel effect between *semiconducting members* for $T=300°$K and $n=10^{16}$. All calculations for this figure are based on Eq. (30).

The maximum near the σ axis (dashed branch) is more pronounced here than in Fig. 6; nevertheless, it will have the same explanation. Therefore, the horizontal beginnings of the curves have been indicated (solid lines). An ohmic region of σ of this order has been stated experimentally by Braun and Busch.[14]

All the curves have been extended to regions of very strong fields F, but it must be remembered that they probably lose their validity for fields of the order 10^7 volt/cm$=0.1$ volt/A or more. This is the order of the limit for the macroscopic electric strength of any material.† The voltages V, which correspond to the field strength $F=0.1$ volt/A, have been indicated by heavy vertical lines in Figs. 6 and 7.

Equation (25) is deduced for the case of $V\geqq\eta+\varphi$, but it still gives a fairly good approximation to the curves of Fig. 6 down to $V=\varphi+1$. However, since even this limit lies beyond the $(F=0.1)$-line in all cases which are represented in Fig. 6, except when $s=60$A, it is obvious that the validity of Eq. (25) is very restricted; a fact to which due attention has not been given by some authors.[15]

Another failure to observe the region of validity of the used equations diminishes the value of Frenkel's results.[16] There is, of course, a theoretical possibility that a current of fast electrons passes through the film on energy levels above the potential hill. This current may be computed by applying Eqs. (34) and (35) of the appendix. The dot-dashed vertical lines in Fig. 6 indicate at what voltage V across the film this current attains the value of 1/10 of the tunnel current. It occurs at $V=5$ to 6 volts, and it is obvious that the voltages which would be able to produce a considerable thermionic current across the film are hardly possible.

APPENDIX

The Richardson-Schottky Effect in Contact Films

The Richardson-Schottky effect is also called thermionic emission and is constituted by the electrons which, because of their thermal movement, are able to overcome the potential hill of the gap. Their amount is approximately expressed[17,18] by the well-known equation

$$j=10^2T^2e^{-11600\phi/T}\ \text{amp/cm}^2, \tag{33}$$

where ϕ electron volts is the top height of the potential hill. With $T=300°$K, Eq. (33) may be written

$$j=9\times 10^6\times 10^{-16.82\phi}\ \text{amp/cm}^2. \tag{34}$$

According to references 1 and 11.

$$\phi=\varphi-V/2-10/s, \tag{35}$$

where V is the voltage applied at the film, and $10/s$ represents the influence of the image force.[12] We have used Eqs. (34) and (35) to calculate the voltage at which the thermionic current from the left number constitutes 10 percent of the field effect with given φ and s. The emission from the right member is obstructed by the applied voltage and may be neglected. The result of these calculations is represented by the dot-dash curves in Fig. 6; see Sec. VII.

[13] R. Holm and W. Meissner, Kontaktwiderstand zwischen Supraleitern und Nichtsupraleitern. Z. Physik **74**, 715 (1932); **86**, 787 (1933).

[14] A. Braun and G. Busch, Helv. Phys. Acta **15**, 571 (1942), Figs. (2) and (3).

† The electric strength may be higher in discrete points of the film; see reference 2, p. 140.

[15] P. F. Soper, Proc. Inst. Elec. Engrs. **96**, 645 (1949).

[16] J. Frenkel, J. Phys., U.S.S.R. **9**, 489 (1945).

[17] W. Schottky, Z. Physik **14**, 63 (1923).

[18] See reference 12, Eq. (23.03).

Summer Program in Nuclear Engineering at North Carolina State College

North Carolina State College offers a full undergraduate course of study in Nuclear Engineering, leading to the degree of Bachelor of Science in Nuclear Engineering, and a graduate program leading to the degree of Master of Science in Nuclear Engineering.[1] About sixty students are enrolled in these curricula. A summer program, consisting of two successive six-week terms, beginning June 11, has been planned by the College. The program has been devised to meet the needs of three groups:

1. Students now enrolled, both undergraduate and graduate, may advance by a full quarter toward their respective degrees.
2. Students desiring to begin work in this field by transferring into the curricula at advanced undergraduate or graduate levels may do so and use the two summer periods either for review in mathematics and physics or for getting on with regular courses. (Around a dozen additional Air Force officers are expected in this category.)
3. Research workers, laboratory technicians, civilian defense personnel, and others desiring short specialized training in Nuclear Instrumentation, Techniques in Radioactivity, etc., may register for specific courses in these subjects available in each of the two six-week terms.

Applicants for summer school courses will be accepted, to the capacity of existing facilities, until registration dates for the two terms, respectively, which are June 11 and July 23. Application blanks, detailed course descriptions, and information on financial arrangements and living quarters may be obtained from the Physics Department, the School of Engineering, or the Registration Office of the College.

[1] Clifford Beck, "Undergraduate curriculum in nuclear engineering at North Carolina State College," *Nucleonics* (January, 1951).

On the Diffraction of a Plane Electromagnetic Wave by a Paraboloid of Revolution

C. W. HORTON AND F. C. KARAL, JR.
The University of Texas, Austin, Texas
(Received September 11, 1950)

The problem of the diffraction of a plane electromagnetic wave by the convex surface of a paraboloid of revolution is examined. Expressions for the components of the incident wave, the scattered wave, and the refracted wave are obtained and are written in such a form that boundary conditions can be satisfied over a paraboloidal surface $\xi=\xi_0$. The analysis is then restricted to the case of a perfectly conducting paraboloid and a wave whose wave front is perpendicular to the axis of rotation. The amplitude of the scattered wave is plotted as a function of the distance along the axis of rotation for three paraboloids. One of these curves is compared with the scattered field produced by a sphere whose radius is equal to the radius of curvature of the paraboloid at the nose.

I. INTRODUCTION

THE first theoretical analysis of the behavior of electromagnetic waves in the presence of a perfectly conducting paraboloid of revolution was made by Abraham.[1] This analysis was undertaken because of some experimental measurements of Birkeland and Sarasin[2] who showed that an electromagnetic wave traveling along a semi-infinite wire was reflected from the free end. Abraham expressed the wave equation in paraboloidal coordinates and obtained series solutions. From these he calculated an approximate expression for the distance between the current nodes in the wire. The resulting values did not agree very well with the experimental results and the problem was dropped.

Except for a brief mention by Lamb,[3] the behavior of electromagnetic waves in the vicinity of a perfectly conducting paraboloid was not considered further until the parabolic reflector was used for Radar waves. Recently Pinney[4,5] and Buchholz[6] have treated various aspects of the problem.

The problem of the diffraction of an incident plane electromagnetic wave by a perfectly conducting paraboloid of revolution is of importance in several problems. For example, the solution to the problem will give detailed information on the field strengths in the neighborhood of the nose of a subsonic missile or of a large bomber; information that will be useful in the location of antennas.

The problem of the diffraction of a plane electromagnetic wave by a perfectly conducting prolate ellipsoid of revolution has been treated recently by Schultz.[7] For many objects the prolate ellipsoid and the paraboloid form equally good approximations. However, if the incident plane wave is perpendicular to the axis of revolution, the use of paraboloidal coordinates offers significant computational advantages. This follows from the fact that in the present analysis the boundary equations which determine the coefficients of the scattered wave can be separated into pairs of equations which can be solved exactly as shown in Eqs. (41) and (42), whereas in the case of ellipsoidal coordinates it is necessary to assume in advance that the higher order coefficients are zero. On the other hand, the functions $V_n^m(2ik\xi)$ have asymptotic forms which cannot be interpreted as a spherical wave multiplied by a function of the direction so that the equivalent cross-section area for scattering cannot be computed easily.

E. Pinney has developed extensively the functions that are necessary in order to discuss electromagnetic waves in paraboloidal coordinates. Since he confined his applications to radiating dipoles located at the focus of a perfectly conducting paraboloid, it is of interest to extend the analysis to the case of a plane wave impinging on the convex surface of a paraboloid.

Pinney, Eq. (5.2), has shown that a general solution of the scalar wave equation

$$\nabla^2\psi+k^2\psi=0 \tag{1}$$

for periodic fields in a homogeneous and isotropic medium may be written as

$$\psi=\{AS_\nu^\mu(\pm 2ik\xi)+BV_\nu^\mu(\pm 2ik\xi)\}\times\{CS_\nu^\mu(\mp 2ik\eta)+DV_\nu^\mu(\mp 2ik\eta)\}\times\{Ee^{i\mu\phi}+Fe^{-i\mu\phi}\}, \tag{2}$$

where ξ, η, ϕ = paraboloidal coordinates illustrated in Fig. 1 and μ, ν = arbitrary real or complex parameters. These solutions may be summed or integrated with respect to μ and ν. It is clear that if ψ is a single-valued function of ϕ, then μ is restricted to integral values of m. The various combinations of signs correspond to waves traveling in different directions. These correspondences are tabulated by Pinney in Eqs. (5.32) and (5.33) for a time factor of the form $\exp(-i\omega t)$ which is used in this paper.

The function $S_\nu^\mu(z)$, defined by Pinney in Eqs. (5.1) and (2.1), is regular at $z=0$, and so it may be used in both the regions inside and outside of any paraboloid.

[1] M. Abraham, Ann. Phys., Lpz. **2**, 32 (1900).
[2] K. Birkeland, and E. Sarasin, Compt. rend. **117**, 618 (1893).
[3] H. Lamb, Proc. Lond. Math. Soc. **4**, 190 (1906).
[4] E. Pinney, J. Math. Phys. **25**, 49 (1946).
[5] E. Pinney, J. Math. Phys. **26**, 42 (1947). Since the equations in these two articles are numbered consecutively, the two articles will not be distinguished in the references.
[6] H. Buchholz, Ann. Phys., Lpz. **2**, 185 (1948).
[7] F. V. Schultz, "Scattering by a prolate spheroid," Report UMM-42, Aeronautical Research Center, Engineering Research Institute, University of Michigan, March 1, 1950.

On the other hand, $V_{\nu}^{\mu}(z)$ has a singularity at $z=0$, and consequently it may be used only in regions that do not include the origin.

Stratton[8] has shown that vector solutions of the vector wave equation,

$$\nabla\nabla\cdot\mathbf{C}-\nabla\times\nabla\times\mathbf{C}+k^2\mathbf{C}=0 \tag{3}$$

can be constructed from scalar solutions of Eq. (1). These solutions are

$$\mathbf{L}=\nabla\psi;\quad \mathbf{M}=\nabla\times\mathbf{a}\psi;\quad \mathbf{N}=k^{-1}\nabla\times\mathbf{M} \tag{4}$$

where $\mathbf{a}$ is a constant unit vector. When there are no "free" sources at a finite distance from the origin, it is not necessary to use the vector $\mathbf{L}$ so that one may build up electromagnetic fields in the form

$$\left.\begin{aligned}\mathbf{E}&=-(A\mathbf{M}+C\mathbf{N})\\ \mathbf{H}&=(ik/\omega\mu)(A\mathbf{N}+C\mathbf{M}).\end{aligned}\right\} \tag{5}$$

The vector $\mathbf{a}$ is chosen as a unit vector parallel to the $+z$-axis. When the vector $\mathbf{M}$ is computed by forming the curl, one obtains, after some manipulation,

$$\left.\begin{aligned}M_{\xi}&=-\frac{1}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\frac{\partial\psi}{\partial\phi}\\ M_{\eta}&=-\frac{1}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\frac{\partial\psi}{\partial\phi}\\ M_{\phi}&=+\frac{(\xi\eta)^{\frac{1}{2}}}{\xi+\eta}\left\{\frac{\partial\psi}{\partial\xi}+\frac{\partial\psi}{\partial\eta}\right\}.\end{aligned}\right\} \tag{6}$$

Further calculations yield the following components for N:

$$\left.\begin{aligned}N_{\xi}&=\frac{1}{k}\left(\frac{\xi}{\xi+\eta}\right)^{\frac{1}{2}}\left[-k^2\psi+\frac{\partial}{\partial\xi}\left\{\frac{-\xi\psi_{\xi}+\eta\psi_{\eta}}{\xi+\eta}\right\}\right]\\ N_{\eta}&=\frac{1}{k}\left(\frac{\eta}{\xi+\eta}\right)^{\frac{1}{2}}\left[+k^2\psi+\frac{\partial}{\partial\eta}\left\{\frac{-\xi\psi_{\xi}+\eta\psi_{\eta}}{\xi+\eta}\right\}\right]\\ N_{\phi}&=\frac{1}{k}\frac{1}{2(\xi\eta)^{\frac{1}{2}}(\xi+\eta)}\frac{\partial}{\partial\phi}\left[\eta\frac{\partial\psi}{\partial\eta}-\xi\frac{\partial\psi}{\partial\xi}\right].\end{aligned}\right\} \tag{7}$$

The vectors $\mathbf{M}$ and $\mathbf{N}$ do not appear to have the necessary orthogonality properties to enable one to expand an arbitrary vector function in terms of them directly. However, it will be seen that the expressions may be arranged so that the orthogonality properties of the S_{ν}^{μ} may be utilized to evaluate coefficients.

II. WAVES TRAVELING IN THE $+z$ DIRECTION

Consider the problem of a plane wave impinging on the paraboloid $\xi=\xi_0$ as shown in Fig. 1. There is no loss in generality in assuming that the direction of wave propagation is in the plane $\phi=0$ and π. This plane wave will produce a scattered or diffracted field in region 1 which may be characterized as an outgoing wave, and a refracted wave in region 2 which may be characterized as a wave traveling in the $+z$ direction.

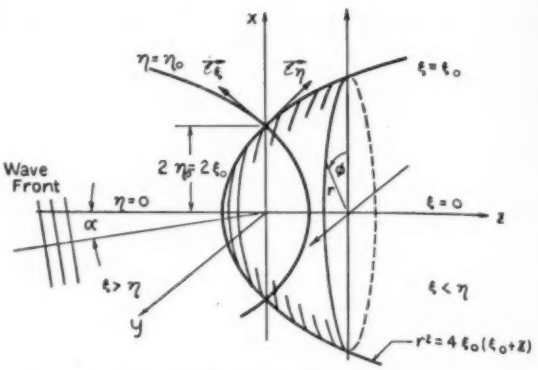


FIG. 1. The paraboloidal coordinate system and the incident plane wave.

Buchholz[9] has shown that a scalar plane wave may be expanded in the following series of paraboloidal functions.

$$\pi=\sum_{m,\,n=0}^{\infty}\theta_{mn}S_n^{m}(+2ik\xi)S_n^{m}(-2ik\eta)\cos m\phi\exp(-i\omega t), \tag{8}$$

where

$$\theta_{mn}=\frac{(-)^n n!}{(m+n)!}\epsilon_m i^m\frac{\tan^{m+2n}(\alpha/2)}{\cos^2(\alpha/2)} \tag{9}$$

and

$$\epsilon_m=\text{Neumann's factor}\begin{cases}=1, & m=0\\ =2, & m\neq 0.\end{cases}$$

The expansion (8) is valid only for $|\alpha|<\pi/2$. This restriction has been removed by Buchholz who replaced the summation over n by an integration over complex values of n.

When the electric field components (5) are computed for the scalar plane wave (8), the electric vector is parallel to the plane of incidence when $A=0$ and perpendicular to the plane of incidence when $C=0$.

It is necessary to develop explicit expressions for the electromagnetic field components for the scalar wave function (henceforth the time factor $\exp(-i\omega t)$ will be omitted in all expressions):

$$\psi_{e_o}=\sum_{m,\,n=0}^{\infty}A_{mn}S_n^{m}(+2ik\xi)S_n^{m}(-2ik\eta)\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \tag{10}$$

The resulting expressions will apply both to the plane wave (8) and to the refracted wave in region 2.

In order to calculate the components of N defined in (7) it is necessary to evaluate

$$(-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta})/(\xi+\eta). \tag{11}$$

[8] J. A. Stratton, *Electromagnetic Theory*, (McGraw-Hill Book Company, Inc., New York, 1941), pp. 392–394. In consequence M.K.S. units are used throughout this paper.

[9] H. Buchholz, Z. Angew. Math. Mech. **23**, 114 (1943).

This combination has been considered by Pinney, Eq. (5.29). After some manipulation one arrives at

$$\frac{-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta}}{\xi+\eta}=2ik\sum_{m,\,n=0}^{\infty}A_{mn}\frac{(m+n)!}{n!}\sum_{p=0}^{n-1}\frac{p!}{(p+m)!}\times S_p^m(-2ik\eta)S_p^m(2ik\xi)\begin{matrix}\cos\\ \sin\end{matrix}m\phi+ik\sum_{m,\,n=0}^{\infty}A_{mn}S_n^m(-2ik\eta)S_n^m(2ik\xi)\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \quad (12)$$

It is obvious that this expression is not in a suitable form for application to boundary conditions at $\xi=\xi_0$. Fortunately, however, the first series on the right side may be rearranged to read

$$2ik\sum_{m,\,n=0}^{\infty}\left(\sum_{p=n+1}^{\infty}\frac{(m+p)!n!}{(m+n)!p!}A_{mp}\right)\times S_n^m(-2ik\eta)S_n^m(+2ik\xi)\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \quad (13)$$

Consequently, one may write

$$\frac{-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta}}{\xi+\eta}=ik\sum_{m,\,n=0}^{\infty}\{2B_{mn}-A_{mn}\}\times S_n^m(-2ik\eta)S_n^m(+2ik\xi)\begin{matrix}\cos\\ \sin\end{matrix}m\phi \quad (14)$$

where

$$B_{mn}=\sum_{p=n}^{\infty}\frac{(m+p)!n!}{(m+n)!p!}A_{mp}. \quad (15)$$

The next step involves a computation which yields

$$k^2\psi_{e_o}+\frac{\partial}{\partial\eta}\left\{\frac{-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta}}{\xi+\eta}\right\}=\frac{ik}{2\eta}\sum_{m,\,n=0}^{\infty}2B_{mn}S_n^m(2ik\xi)\{(n+1)S^m{}_{n+1}(-2ik\eta)-(m+n)S^m{}_{n-1}(-2ik\eta)-S_n^m(-2ik\eta)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi+\frac{ik}{2\eta}\sum_{m,\,n=0}^{\infty}A_{mn}S_n^m(2ik\xi)\{-2(n+1)S^m{}_{n+1}(-2ik\eta)+(m+2n+2)S_n^m(-2ik\eta)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \quad (16)$$

Since this series is not suitable for boundary conditions involving $\xi=\xi_0$, it must be rearranged. This yields

$$k^2\psi_{e_o}+\frac{\partial}{\partial\eta}\left\{\frac{-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta}}{\xi+\eta}\right\}=\frac{ik}{2\eta}\sum_{m,\,n=0}^{\infty}2S_n^m(-2ik\eta)\times\{nB_{m,\,n-1}S^m{}_{n-1}(2ik\xi)-B_{mn}S_n^m(2ik\xi)-(m+n+1)B_{m,\,n+1}S^m{}_{n+1}(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi+\frac{ik}{2\eta}\sum_{m,\,n=0}^{\infty}S_n^m(-2ik\eta)\{-2nA_{m,\,n-1}S^m{}_{n-1}(2ik\xi)+(m+2n+2)A_{mn}S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \quad (17)$$

A similar extensive calculation yields

$$-k^2\psi_{e_o}+\frac{\partial}{\partial\xi}\left\{\frac{-\xi\psi_{e_o\xi}+\eta\psi_{e_o\eta}}{\xi+\eta}\right\}=\frac{ik}{2\xi}\sum_{m,\,n=0}^{\infty}2B_{mn}S_n^m(-2ik\eta)\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)-S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi+\frac{ik}{2\xi}\sum_{m,\,n=0}^{\infty}A_{mn}S_n^m(-2ik\eta)\{-2(n+1)S^m{}_{n+1}(2ik\xi)+(m+2n+2)S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix}m\phi. \quad (18)$$

It is evident from (5) that two distinct solutions can be constructed from two different scalar functions ψ and ψ^*. If in parallelism with (10) one writes

$$\psi_{e_o}{}^*=\sum_{m,\,n=0}^{\infty}C_{mn}S_n^m(+2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\ \sin\end{matrix}m\phi; \quad (19)$$

then Eqs. (12)–(18) may be repeated with ψ replaced by ψ^*, A_{mn} replaced by C_{mn}, and B_{mn} replaced by D_{mn} which is defined by

$$D_{mn}=\sum_{p=n}^{\infty}\frac{(m+p)!n!}{(m+n)!p!}C_{mp}. \quad (20)$$

Equations (10)–(18) for ψ and their counterpart for ψ^* enable one to write out the components for E_ξ, E_η, H_ξ, and H_η. This is done in Eqs. (25), (26), (28), and (29). In order to calculate E_ϕ and H_ϕ it is necessary to evaluate

$$\frac{(\xi\eta)^{\frac{1}{2}}}{\xi+\eta}(\psi_{e_o\xi}+\psi_{e_o\eta}). \quad (21)$$

After some manipulation and a rearrangement of series one obtains

$$\frac{(\xi\eta)^{\frac{1}{2}}}{\xi+\eta}(\psi_{e_{o\xi}}+\psi_{e_{o\eta}})=\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}S_n^m(-2ik\eta)[nF_{mn}\{nS_n^m(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)\}$$

$$-(m+n+1)F_{m,n+1}\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n+1)S_n^m(2ik\xi)\}]\begin{matrix}\cos\\ \sin\end{matrix} m\phi$$

$$+\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}\frac{m}{2}A_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\ \sin\end{matrix} m\phi \quad (22)$$

where

$$F_{mn}=\sum_{p=n}^{\infty}\frac{(m+p)!(n-1)!}{(m+n)!p!}A_{mp}. \quad (23)$$

Similarly, if ψ is replaced by ψ^* in (22), one should replace A_{mn} by C_{mn} and F_{mn} by

$$G_{mn}=\sum_{p=n}^{\infty}\frac{(m+p)!(n-1)!}{(m+n)!p!}C_{mp}. \quad (24)$$

F_{mo} and G_{mo} are defined to be zero.

The field components of $\mathbf{E}$ and $\mathbf{H}$ can be evaluated from Eq. (5) with the aid of the succeeding expressions. One obtains

$$E_\xi=\pm\frac{1}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}mA_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\ \sin\end{matrix} m\phi$$

$$-\frac{i}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}2D_{mn}S_n^m(-2ik\eta)\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)-S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix} m\phi$$

$$-\frac{i}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}C_{mn}S_n^m(-2ik\eta)\{-2(n+1)S^m{}_{n+1}(2ik\xi)+(m+2n+2)S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix} m\phi. \quad (25)$$

$$E_\eta=\pm\frac{1}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}mA_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\ \sin\end{matrix} m\phi$$

$$-\frac{i}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}2S_n^m(-2ik\eta)\{nD_{m,n-1}S^m{}_{n-1}(2ik\xi)-D_{mn}S_n^m(2ik\xi)-(m+n+1)D_{m,n+1}S^m{}_{n+1}(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix} m\phi$$

$$-\frac{i}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}S_n^m(-2ik\eta)\{-2nC_{m,n-1}S^m{}_{n-1}(2ik\xi)+(m+2n+2)C_{mn}S_n^m(2ik\xi)\}\begin{matrix}\cos\\ \sin\end{matrix} m\phi. \quad (26)$$

$$E_\phi=-\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}(m/2)A_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\sin\\ \cos\end{matrix} m\phi$$

$$-\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}S_n^m(-2ik\eta)[nF_{mn}\{nS_n^m(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)\}$$

$$-(m+n+1)F_{m,n+1}\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n+1)S_n^m(2ik\xi)\}]\begin{matrix}\sin\\ \cos\end{matrix} m\phi$$

$$\pm\frac{i}{2(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}m\{2D_{mn}-C_{mn}\}S_n^m(-2ik\eta)S_n^m(2ik\xi)\begin{matrix}\sin\\ \cos\end{matrix} m\phi. \quad (27)$$

$$H_\xi=-\frac{k}{\omega\mu}\frac{1}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}2B_{mn}S_n^m(-2ik\eta)\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)-S_n^m(2ik\xi)\}\begin{matrix}\sin\\\cos\end{matrix}m\phi$$

$$-\frac{k}{\omega\mu}\frac{1}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}A_{mn}S_n^m(-2ik\eta)\{-2(n+1)S^m{}_{n+1}(2ik\xi)+(m+2n+2)S_n^m(2ik\xi)\}\begin{matrix}\sin\\\cos\end{matrix}m\phi$$

$$\pm\frac{ik}{\omega\mu}\frac{1}{2[\xi(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}mC_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\sin\\\cos\end{matrix}m\phi. \tag{28}$$

(22)

$$H_\eta=-\frac{k}{\omega\mu}\frac{1}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}2S_n^m(-2ik\eta)\{nB_{m,n-1}S^m{}_{n-1}(2ik\xi)$$

(23)

$$-B_{mn}S_n^m(2ik\xi)-(m+n+1)B_{m,n+1}S^m{}_{n+1}(2ik\xi)\}\begin{matrix}\sin\\\cos\end{matrix}m\phi$$

$$-\frac{k}{\omega\mu}\frac{1}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}S_n^m(-2ik\eta)\{-2nA_{m,n-1}S^m{}_{n-1}(2ik\xi)+(m+2n+2)A_{mn}S_n^m(2ik\xi)\}\begin{matrix}\sin\\\cos\end{matrix}m\phi$$

(24)

$$\pm\frac{ik}{\omega\mu}\frac{1}{2[\eta(\xi+\eta)]^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}mC_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\sin\\\cos\end{matrix}m\phi. \tag{29}$$

s. One

$$H_\phi=\mp\frac{k}{2\omega\mu}\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}m\{2B_{mn}-A_{mn}\}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\\sin\end{matrix}m\phi$$

$$+\frac{ik}{\omega\mu}\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}S_n^m(-2ik\eta)[nG_{mn}\{nS_n^m(2ik\xi)-(m+n)S^m{}_{n-1}(2ik\xi)\}$$

(25)

$$-(m+n+1)G_{m,n+1}\{(n+1)S^m{}_{n+1}(2ik\xi)-(m+n+1)S_n^m(2ik\xi)\}]\begin{matrix}\cos\\\sin\end{matrix}m\phi$$

$$+\frac{ik}{\omega\mu}\frac{1}{(\xi\eta)^{\frac{1}{2}}}\sum_{m,n=0}^{\infty}(m/2)C_{mn}S_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\\sin\end{matrix}m\phi. \tag{30}$$

cos
$m\phi$
sin

(26)

It will be noticed that when $\xi=\xi_0$, a constant, all of the field components are of the form,

$$\sum_{m,n=0}^{\infty}a_{mn}S_n^m(-2ik\eta)\begin{matrix}\cos\\\sin\end{matrix}m\phi.$$

Since these two sets of functions are orthogonal[10] and, presumably, complete over the surface of the paraboloid, the coefficients A_{mn} and C_{mn} can be determined so that $E_{\tan}$ will take on any prescribed value over $\xi=\xi_0$.

The field components (25)–(30) represent waves traveling in the $+z$ direction. When $k=k_2$, they can be used to represent the fields produced in region 2 by the plane wave. Further, when $k=k_1$, the components represent an incident plane wave in accordance with the following scheme:

A. Electric vector parallel to the plane of incidence:

$$A_{mn}=0 \quad \text{and} \quad C_{mn}=E_0\theta_{mn}.$$

(27)

B. Electric vector perpendicular to the plane of incidence:

$$A_{mn}=E_0\theta_{mn} \quad \text{and} \quad C_{mn}=0.$$

III. THE SCATTERED WAVE

In addition to the incident plane wave in region 1 and the refracted wave in region 2, there is a scattered wave in region 1. In order to satisfy boundary conditions over $\xi=\xi_0$, it is necessary that the scattered wave depend on $S_n^m(-2ik\eta)$. In order that the expansion represent an outgoing wave, the dependence on ξ must be of the form $V_n^m(2ik\xi)$. Thus one must consider scalar wave functions of the form,

$$\psi_{eo}=\sum_{m,n=0}^{\infty}\mathfrak{a}_{mn}V_n^m(2ik\xi)S_n^m(-2ik\eta)\begin{matrix}\cos\\\sin\end{matrix}m\phi. \tag{31}$$

An analogous wave function ψ^* will be considered in which $\mathfrak{a}_{mn}$ is replaced by $\mathfrak{e}_{mn}$. After a lengthy manipulation similar to that outlined above, one can obtain

[10] See reference 4, Eq. (2.26).

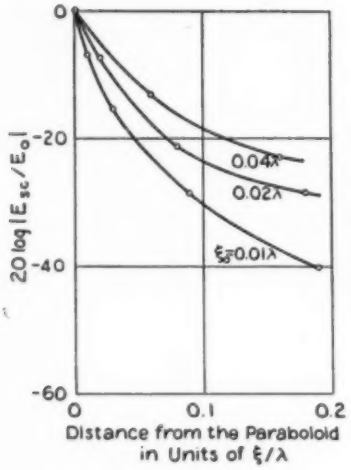


FIG. 2. The amplitude of the scattered wave produced by a perfectly conducting paraboloid whose axis of rotation is normal to the incident wave front.

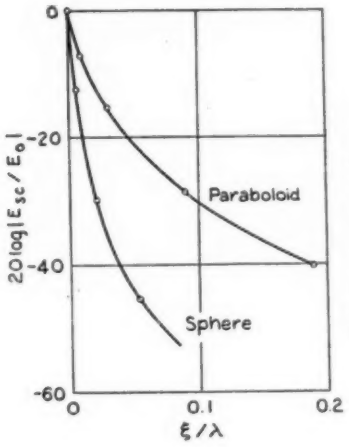


FIG. 3. A comparison of the amplitudes of the scattered waves produced by a perfectly conducting paraboloid and by a perfectly conducting sphere of the same radius of curvature. Radius of the sphere is $0.02\xi_0$ and the paraboloid is $\xi_0=0.01$.

explicit expressions for the field components. These components are identical in form with Eqs. (25)–(30). The only changes necessary are that $S_n^m(2ik\xi)$ should be replaced everywhere by $V_n^m(2ik\xi)$, and all block capital letters should be replaced by script letters. The script letters are defined by

$$\mathfrak{B}_{mn}=-\sum_{p=0}^{n-1}\frac{(m+p)!n!}{(m+n)!p!}\mathfrak{a}_{mp} \tag{32}$$

$$\mathfrak{D}_{mn}=-\sum_{p=0}^{n-1}\frac{(m+p)!n!}{(m+n)!p!}\mathfrak{c}_{mp} \tag{33}$$

$$\mathfrak{F}_{mn}=-\sum_{p=0}^{n-1}\frac{(m+p)!(n-1)!}{(m+n)!p!}\mathfrak{a}_{mp} \tag{34}$$

$$\mathfrak{G}_{mn}=-\sum_{p=0}^{n-1}\frac{(m+p)!(n-1)!}{(m+n)!p!}\mathfrak{c}_{mp}. \tag{35}$$

These coefficients are zero when $n=0$. $\mathfrak{a}_{mp}$ and $\mathfrak{c}_{mp}$ are of course the fundamental coefficients that are to be determined by the boundary conditions.

IV. BOUNDARY CONDITIONS

A formal solution to the problem of the scattering of a plane wave by a paraboloid of arbitrary material can be obtained by equating the tangential components of **E** and **H**. Actually the formulas are so complicated that nothing short of a large computing program would justify the expansion of the equations. The problem of the scattering of an arbitrary plane wave by a perfectly conducting paraboloid is quite manageable, and this problem will be returned to in a future paper after numerical values of the paraboloidal functions are available.

V. THE CASE $\alpha=0$ FOR THE PERFECTLY CONDUCTING PARABOLOID

A plane wave whose wave front is perpendicular to the z-axis can be readily expressed in paraboloidal functions without recourse to Buchholz's expansion (8). By definition

$$S_0^1(z)=z^{\frac{1}{2}}\exp(-z/2) \tag{36}$$

so that

$$\exp(ikz)=\exp\{ik(\eta-\xi)\}$$
$$=\frac{1}{2k(\xi\eta)^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta). \tag{37}$$

Pinney, Eq. (7.3), has shown that the unit cartesian vector $\mathbf{i}_x$ is related to the paraboloidal unit vectors by

$$\mathbf{i}_x=\left(\frac{\eta}{\xi+\eta}\right)^{\frac{1}{2}}\cos\phi\mathbf{i}_\xi+\left(\frac{\xi}{\xi+\eta}\right)^{\frac{1}{2}}\cos\phi\mathbf{i}_\eta-\sin\phi\mathbf{i}_\phi. \tag{38}$$

Consequently, an incident plane wave which is polarized with the electric vector parallel to the x-axis may be written as

$$\left.\begin{aligned}
E_\xi^{(in)}&=\frac{E_0}{2k}\frac{1}{[\xi(\xi+\eta)]^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\cos\phi\\
E_\eta^{(in)}&=\frac{E_0}{2k}\frac{1}{[\eta(\xi+\eta)]^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\cos\phi\\
E_\phi^{(in)}&=-\frac{E_0}{2k}\frac{1}{(\xi\eta)^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\sin\phi.
\end{aligned}\right\} \tag{39}$$

The corresponding components of **H** are

$$\left.\begin{aligned}
H_\xi^{(in)}&=\frac{E_0}{2\mu\omega}\frac{1}{[\xi(\xi+\eta)]^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\sin\phi\\
H_\eta^{(in)}&=\frac{E_0}{2\mu\omega}\frac{1}{[\eta(\xi+\eta)]^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\sin\phi\\
H_\phi^{(in)}&=\frac{E_0}{2\mu\omega}\frac{1}{(\xi\eta)^{\frac{1}{2}}}S_0^1(2ik\xi)S_0^1(-2ik\eta)\cos\phi.
\end{aligned}\right\} \tag{40}$$

The sum of the incident electric field $\mathbf{E}^{(in)}$ and of the scattered field should be such that the tangential components E_η and E_ϕ vanish at $\xi=\xi_0$. Equation (39) shows

that the only terms present in the scattered field are those for $m=1$. The resulting pair of equations are each summed over n, and the terms have $S_n^1(-2ik\eta)$ as a common factor. Since these functions are orthogonal, it is necessary that the coefficients vanish separately. These considerations lead to an infinite set of pairs of equations which may be solved for $\mathfrak{a}_{1n}$ and $\mathfrak{c}_{1n}$.

For $n=0$ one has (the argument $+2ik\xi_0$ is omitted)

$$\left.\begin{aligned} V_0^1\mathfrak{a}_{10} \quad & -i(3V_0^1+2V_1^1)\mathfrak{c}_{10} = -(E_0/k)S_0^1 \\ (3V_0^1-2V_1^1)\mathfrak{a}_{10} \quad & -iV_0^1\mathfrak{c}_{10} = +(E_0/k)S_0^1. \end{aligned}\right\} \tag{41}$$

For $n=1$ one has

$$\left.\begin{aligned} V_1^1\mathfrak{a}_{11} \quad & -i(5V_1^1+4V_2^1)\mathfrak{c}_{11} = -i(2V_0^1-V_1^1-2V_2^1)\mathfrak{c}_{10} \\ (5V_1^1-4V_2^1)\mathfrak{a}_{11} \quad & -iV_1^1\mathfrak{c}_{11} = (2V_0^1-4V_1^1+2V_2^1)\mathfrak{a}_{10}+iV_1^1\mathfrak{c}_{10}, \end{aligned}\right\} \tag{42}$$

etc. Hence one is lead to a sequence of equations that can be solved to any degree of accuracy that is desired.

Since no numerical values of the functions S_n^m and V_n^m are available, only a limited number of computations have been made. Figure 2 shows the ratio $E_\eta(\text{scattered})/E_0$ for points along the axis of rotation for three paraboloids characterized by $\xi_0/\lambda=0.01, 0.02, 0.04$. The computation has been carried only to the terms in $n=0$, but this approximation appears to be accurate within a few percent, since the ratios $E_\eta(\text{scattered})/E_0$ at the surface of the paraboloid are 1.06, 1.08, 1.08 for $\xi_0/\lambda=0.01, 0.02, 0.04$, respectively.

For the purpose of comparison, the scattered field due to a perfectly conducting sphere whose radius is equal to the radius of curvature of the paraboloid has been computed for the paraboloid $\xi_0/\lambda=0.01$, i.e., for the sphere whose radius is $0.02\xi_0$. Figure 3 shows the scattered fields of the sphere and the paraboloid.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

A Problem in Heat Conduction

BORIS PODOLSKY
Physics Department, University of Cincinnati, Cincinnati, Ohio
(Received September 27, 1950)

A steady-state temperature distribution is found for the case of a periodically varying point source of heat, immersed in a homogeneous isotropic infinitely extended fluid, moving with a constant velocity. The phase velocity of the generated temperature wave is discussed. If the phase velocity in a stationary medium is V_0, and the fluid velocity $v>V_0$, the phase velocity in a moving medium is given by $V=v[1+\frac{1}{2}(V_0/v)^4-(5/8)(V_0/v)^8+\cdots]$. For practical cases one can usually assume $V=v$, with an error of less than 0.1 percent, even when V_0 is as large as 20 percent of v.

THE problem discussed in this paper arises out of an idealization of the following practical problem: A small coil of resistance wire, immersed in a moving fluid, is supplied with alternating current of frequency f, which delivers to the coil an average power of $\bar{W}$ calories per second. What will be the temperature distribution within the fluid? The problem is idealized in the following ways: (1) The coil is replaced by a fixed point source of heat, which delivers to the fluid sinusoidally varying power (in calories per second)

$$W=\bar{W}(1-\cos\omega t) \tag{1}$$

where $\omega=4\pi f$. This is, of course, the way in which power produced by an alternating current would vary with time. (2) The medium is assumed to be homogeneous, isotropic, and infinitely extended. (3) The medium is assumed to move with uniform speed v in the direction of the Z-axis. (4) The time t, at which the temperature distribution is sought, is assumed to be sufficiently long after the start of the experiment to assure the possibility of neglecting transient effects. (5) The initial temperature of the medium, as well as the temperature infinitely far from the source at all times, is the ambient temperature T_a.

The usual heat conduction equation,

$$\nabla^2T=\partial T/\kappa\partial t \tag{2}$$

is assumed to hold in a coordinate system moving with the fluid. Here

$$\kappa=k/\rho C, \tag{3}$$

where k, ρ, and C are, respectively, the conductivity, the density, and the heat capacity per unit mass of the fluid medium. When transformed nonrelativistically to a coordinate system fixed with respect to the source,

Eq. (2) becomes:

$$\nabla^2 T=\frac{1}{\kappa}\left(\frac{\partial T}{\partial t}+v\frac{\partial T}{\partial z}\right). \tag{4}$$

Equations (2) and (4) merely express the law of conservation of heat in the medium. A point source of heat, where heat is created, is a point at which these equations cannot hold. It must therefore be a singular point of the solution of Eqs. (2) and (4). Before attempting to solve either of these equations, we must investigate the nature of this singularity. We do this by first solving the problem for a spherical source in a stationary medium. This problem is comparatively easy, and we obtain, with neglect of transient effects,

$$T=\frac{\bar{W}}{4\pi kr}\left\{1-\frac{e^{-\alpha(r-a)}}{(1+2\alpha a+2\alpha^2a^2)^{\frac{1}{2}}}\times\cos[\omega t-\alpha(r-a)-\psi]\right\}+T_a \tag{5}$$

where a is the radius of the sphere,

$$\alpha=(\omega/2\kappa)^{\frac{1}{2}} \quad \text{and} \quad \psi=\tan^{-1}(\alpha a/1+\alpha a). \tag{6}$$

The temperature T is here of course a function of time t and of the distance r from the source.

The solution (5) satisfies Eq. (2) for $r>a$, as can be easily verified by substitution. Further, for $r\to\infty$, $T\to T_a$, the ambient temperature. The heat-current density vector, giving in magnitude and direction of the heat flow per unit area per unit of time, is

$$\mathbf{j}=-k\nabla T=-(k\mathbf{r}/r)(\partial T/\partial r), \tag{7}$$

where $\mathbf{r}$ is the radius vector to the point considered.

The total heat flow per second through a surface of radius r is now

$$W=-4\pi kr^2\partial T/\partial r$$

$$=\bar{W}\left\{1-\left(\frac{1+2\alpha r+2\alpha^2r^2}{1+2\alpha a+2\alpha^2a^2}\right)^{\frac{1}{2}}e^{-\alpha(r-a)}\cos(\omega t-\phi)\right\} \tag{8}$$

where

$$\phi=\alpha(r-a)+\tan^{-1}\left(\frac{a\alpha}{1+\alpha a}\right)-\tan^{-1}\left(\frac{\alpha r}{1+\alpha r}\right). \tag{9}$$

For $r=a$, this becomes $W=\bar{W}(1-\cos\omega t)$ in agreement with Eq. (1); but also, for all r, the time average of W is $\bar{W}$, since the average of $\cos(\omega t-\phi)$ is zero. Thus, expression (5) satisfies all the requirements.

From Eq. (5) we obtain the desired solution for a point source in a stationary medium by putting $a=0$. We thus obtain:

$$T=(A/r)\{1-e^{-\alpha r}\cos(\omega t-\alpha r)\}+T_a, \tag{10}$$

where $A=\bar{W}/4\pi k$, and $\alpha=(\omega/2\kappa)^{\frac{1}{2}}$. For $r\to 0$, $T\to\infty$ as $A(1-\cos\omega t)/r$. As was expected the source, $r=0$, is a singular point of the solution.

Our problem may now be stated thus: Find that solution of Eq. (4) which for $v\to 0$ approaches the expression (10).

We write Eq. (10) in the form,

$$T=T_a-A(T_{01}-T_{02}), \tag{11}$$

where

$$T_{01}=(1/r)\exp(-\alpha r)\cos(\omega t-\alpha r) \tag{12}$$

and

$$T_{02}=1/r. \tag{13}$$

We now observe that T_a, T_{01}, and T_{02} each satisfies Eq. (2). Thus, our problem may be divided into three simpler problems, the problems of finding solutions of Eq. (4) which approach T_a, T_{01}, and T_{02}, respectively. Further, T_a, being a constant, satisfies both Eq. (1) and Eq. (4); it can thus be taken over directly, and one of the three problems is solved. We also observe that, for $\omega\to 0$, α approaches zero, so that

$$\lim_{\omega\to 0} T_{01}=T_{02}. \tag{14}$$

Therefore, if we find a solution of Eq. (4) which approaches T_{01} for $v\to 0$, we can expect that putting $\omega=0$ in this solution will give us a solution approaching T_{02} for $v\to 0$. Finally, we note that

$$T_{01}=Re\left(\frac{1}{r}e^{-\alpha r}e^{i(\omega t-\alpha r)}\right)=Re(T_{00}e^{i\omega t}), \tag{15}$$

where $Re(\cdots)$ means the real part of the expression $(\cdots)$, and

$$T_{00}=\frac{1}{r}e^{-\alpha r(1+i)}. \tag{16}$$

$T_{00}e^{i\omega t}$ satisfies Eq. (2), so that if we find a solution of Eq. (4) which approaches this expression for $v\to 0$, its real part will approach T_1.

Equation (4) is a linear equation in T, and the physical system it represents is subject to a periodic forcing of the source. It will therefore have a periodic response with the same angular frequency ω. Thus the solution of Eq. (4) approaching $T_{00}e^{i\omega t}$ as $v\to 0$ will be of the form $T_0e^{i\omega t}$, where $T_0\to T_{00}$ as $v\to 0$.

Our problem is now reduced to the following: Find $T_0(\mathbf{r}, t; v, \omega)$ such that the complex expression,

$$T_c=T_0(\mathbf{r}, t; v, \omega)e^{i\omega t} \tag{17}$$

satisfies Eq. (4) and such that the limit of T_0 for $v\to 0$ is equal to T_{00}, or

$$T_0(\mathbf{r}, t; 0, \omega)=T_{00}=(1/r)e^{-\alpha r(1+i)}. \tag{18}$$

Then the desired solution of our problem will be

$$T=T_a-A(T_1-T_2), \tag{19}$$

where

$$T_1=Re(T_c) \tag{20}$$

and

$$T_2=\lim_{\omega\to 0} T_1. \tag{21}$$

Substituting expression (17) for T into Eq. (4) we obtain:

$$\nabla^2 T_0=-\frac{1}{\kappa}\left(i\omega T_0+v\frac{\partial T_0}{\partial z}\right). \tag{22}$$

We try to solve this equation by the usual method of separation of variables. To do this we must first choose a suitable coordinate system. Assuming the point source to be at the origin, one such a system is the parabolic coordinate system (ξ, η, θ) defined by the relations: (for $\xi\geqslant 0$, $\eta\geqslant 0$)

$$x=(\xi\eta)^{\frac{1}{2}}\cos\theta, \quad y=(\xi\eta)^{\frac{1}{2}}\sin\theta, \quad z=\tfrac{1}{2}(\xi-\eta) \tag{23}$$

from which we can easily work out

$$r=(x^2+y^2+z^2)^{\frac{1}{2}}=\tfrac{1}{2}(\xi+\eta). \tag{24}$$

Since in our problem T_0 must be symmetric about the Z-axis, $\partial T_0/\partial\theta=0$, and therefore Eq. (22) can be written as follows:

$$\frac{4}{\xi+\eta}\left\{\frac{\partial}{\partial\xi}\left(\xi\frac{\partial T_0}{\partial\xi}\right)+\frac{\partial}{\partial\eta}\left(\eta\frac{\partial T_0}{\partial\eta}\right)\right\} =-\frac{1}{\kappa}\left\{i\omega T_0+\frac{2v}{\xi+\eta}\left(\xi\frac{\partial T_0}{\partial\xi}-\eta\frac{\partial T_0}{\partial\eta}\right)\right\}. \tag{25}$$

The method of separation of variables consists in assuming that the desired function $T_0(\xi, \eta)$ can be expressed as a linear combination of products of the type $X(\xi)Y(\eta)$, each of which satisfies Eq. (25). If in Eq. (25) we put $T_0=X(\xi)Y(\eta)$ and divide by $X(\xi)Y(\eta)$ and multiply by $(\xi+\eta)$, the resulting equation is

$$\frac{4}{X}\frac{d}{d\xi}\left(\xi\frac{dX}{d\xi}\right)-\frac{2v}{\kappa}\frac{\xi}{X}\frac{dX}{d\xi}-\frac{i\omega}{\kappa}\xi +\frac{4}{Y}\frac{d}{d\eta}\left(\eta\frac{dY}{d\eta}\right)+\frac{2v}{\kappa}\frac{\eta}{Y}\frac{dY}{d\eta}-\frac{i\omega}{\kappa}\eta=0. \tag{26}$$

The first three terms of this equation, since $X(\xi)$ depends on ξ only, combine into a function of ξ only; that is their sum is independent of η. But this sum, by Eq. (26) is equal to the negative of the sum of the last three terms, which is independent of ξ. Thus, the sum of the first three terms must be a constant, say, b. The sum of the last three terms must then be $-b$. Using this information, Eq. (26) may be split into two:

$$\xi X_b''+\left(1-\frac{v}{2\kappa}\xi\right)X_b'-\frac{1}{4}\left(b+\frac{i\omega}{\kappa}\xi\right)X_b=0 \tag{27}$$

and

$$\eta Y_b''+\left(1+\frac{v}{2\kappa}\eta\right)Y_b'+\frac{1}{4}\left(b-\frac{i\omega}{\kappa}\eta\right)Y_b=0, \tag{28}$$

where $X'=dX/d\xi$, $Y'=dY/d\eta$, etc. The subscript b was attached to X and Y to indicate that the solutions we are considering depend on the choice of the value of b, the so-called *separation constant*.

For each value of b each of the Eqs. (27) and (28) will have two linearly independent solutions, since each is a second-order equation. In general, one of the two solutions will become infinite for infinitely removed points ($\xi\to\infty$ or $\eta\to\infty$). These solutions cannot be admitted, since we wish $T_0\to 0$ at infinity, as it expresses the effect of the source at infinity, where the temperature remains T_a. The other pair of solutions (X_b and Y_b) in general will be well behaved only if certain conditions are imposed on the values of b. We will call values of b satisfying these conditions the *permissible values*. When these values are found, and the corresponding functions X_b and Y_b are calculated, then the expression,

$$T_0=\sum_b a_b X_b(\xi)Y_b(\eta), \tag{29}$$

is the general form of permissible T_0 obtainable by the method of separation of variables. In Eq. (29) the summation is taken over all permissible values of b, a_b's are arbitrary coefficients. If permissible values of b include some continuous range, an integral over this range must be added to the sum in Eq. (29). This possibility must be understood in all that follows.

Our problem would then be solved if we could find the coefficients a_b such that

$$\lim_{v\to 0}\sum_b a_b X_b(\xi)Y_b(\eta)=(1/r)e^{-\alpha r(1+i)}, \tag{30}$$

in accordance with Eqs. (18) and (29). The possibility of such an expansion in a series of products depends upon the fact that $\lim_{v\to 0}X_b(\xi)Y_b(\eta)$ and the right-hand side of Eq. (30) both satisfy Eq. (26) with $v=0$. A direct attack on this problem is, however, very difficult. In fact, this paper owes its existence to certain unexpected simplifications found by the author.

Equations (27) and (28) can be simplified by substitutions*

$$X_b=\exp(-\beta\xi)P_b(\xi) \tag{31}$$

and

$$Y_b=\exp(-\gamma\eta)Q_b(\eta), \tag{32}$$

where β and γ are suitably chosen. The choice is made in such a way that $P_b(\xi)$ and $Q_b(\eta)$ satisfy differential equations of the hypergeometric type. These are equations in which appear but two different powers of the independent variable, when differentiation with respect to this variable is counted as -1st power. Thus, $\xi X_b''$ is of $+1-2=-1$st power, $\xi X_b'$ of power zero, ξX_b of first power. Thus Eq. (27) is not of the hypergeometric type. The simplicity of the hypergeometric type appears when one tries to solve the differential equation by expansion of the unknown function in power series. One then obtains recursion formulas for the coefficients

* $P_b(\xi)$ and $Q_b(\eta)$ are defined by Eqs. (31) and (32), and are not to be confused with the Legendre functions.

of the power series, which for the hypergeometric type are very simple, each relating but two coefficients.

We choose β and γ in such a way as to eliminate terms in ξP_b and ηQ_b. We thus obtain:

$$\xi P_b''+(1-\lambda\xi/2)P_b'-\tfrac{1}{4}(\lambda+\mu)P_b=0 \tag{33}$$

and

$$\eta Q_b''+(1-\lambda\eta/2)Q_b'-\tfrac{1}{4}(\lambda-\mu)Q_b=0, \tag{34}$$

which are of hypergeometric type. Here

$$\lambda=\frac{v}{\kappa}\left[\left(\frac{R+1}{2}\right)^{\frac{1}{2}}+i\left(\frac{R-1}{2}\right)^{\frac{1}{2}}\right], \tag{35}$$

$$R=(1+16\omega^2\kappa^2/v^4)^{\frac{1}{2}}, \tag{36}$$

$$\mu=b-v/\kappa, \tag{37}$$

$$\beta=\tfrac{1}{4}(\lambda-v/\kappa), \tag{38}$$

and

$$\gamma=\tfrac{1}{4}(\lambda+v/\kappa). \tag{39}$$

When $v\to 0$, $\lambda\to\lambda_0=2\alpha(1+i)$, $\mu\to\mu_0=b$, while

$$\beta\to\gamma\to\tfrac{1}{4}\lambda\to\alpha(1+i)/2. \tag{40}$$

Equations (33) and (34) then become:

$$\xi P_b''+(1-\lambda_0\xi/2)P_b'-\tfrac{1}{4}(\lambda_0+\mu_0)P_b=0 \tag{41}$$

and

$$\eta Q_b''+(1-\lambda_0\eta/2)Q_b'-\tfrac{1}{4}(\lambda_0-\mu_0)Q_b=0, \tag{42}$$

which are of the *same form* as Eqs. (33) and (34), only with the parameters λ_0 and μ_0 instead of λ and μ.

We note also that β and γ do not contain b. Therefore, when substitutions (31) and (32) are made in Eq. (29), the exponential factors can be taken outside the summation sign, and we get

$$T_0=e^{-(\beta\xi+\gamma\eta)}\textstyle\sum_b a_b P_b(\xi)Q_b(\eta). \tag{43}$$

Now

$$\lim_{v\to 0}(\beta\xi+\gamma\eta)=\frac{\alpha}{2}(1+i)(\xi+\eta), \text{ by (40)};$$

$$=\alpha r(1+i), \text{ by (24)}.$$

Therefore:

$$\lim_{v\to 0}T_0=\lim_{v\to 0}\textstyle\sum_b a_b X_b(\xi)Y_b(\eta)$$

$$=\exp[-\alpha r(1+i)]\lim_{v\to 0}\textstyle\sum_b a_b P_b(\xi)Q_b(\eta), \text{ by (43)};$$

$$=\frac{1}{r}e^{-\alpha r(1+i)}, \text{ by (30)}.$$

Comparing the last two expressions, we see that we must have:

$$\lim_{v\to 0}\textstyle\sum_b a_b P_b(\xi)Q_b(\eta)=1/r. \tag{44}$$

Thus, to solve our problem, it appears that we still have to do the following:

1. Determine the permissible values of μ_0 and find functions P_b and Q_b satisfying Eqs. (41) and (42).
2. Expand $1/r$ in the series of products of these functions, thus determining the coefficients a_b.
3. Use these coefficients in Eq. (43) with functions P_b and Q_b determined from Eqs. (33) and (34).

Actually a little consideration shows how the desired result can be achieved without any further work.†

$\mu_0=b$ in Eqs. (41) and (42) are the permissible values. These values depend upon the form of the equations (41) and (42) and the value of λ_0, the only other parameter in the equations. Thus, in the left-hand side of Eq. (44), while it is a function of ξ and η, each term will depend upon the value of the parameter λ_0.

In the formation of the sum in Eq. (43) λ_0 is replaced by λ and μ_0 by μ, but the form of the equations is unchanged. Thus, μ will be the same function of λ that μ_0 is of λ_0, and therefore the whole summation will be the same function of λ that the summation in Eq. (44) is of λ_0. But, Eq. (44) must hold for all λ_0. Therefore, when the summation is performed, λ_0 must cancel out.

Thus, the summation in Eq. (43) does not contain λ and is equal to $1/r$. In other words:

$$T_0=\frac{1}{r}e^{-(\beta\xi+\gamma\eta)}=\frac{2}{(\xi+\eta)}e^{-(\beta\xi+\gamma\eta)}. \tag{45}$$

This surprisingly simple result can be immediately verified by substitution of expression (45) into Eq. (25).

Following the prescription of Eqs. (17)–(21) it is now a matter of mere algebra to find the final desired result:

$$T=\frac{A}{r}\exp[-(r-z)v/2\kappa]$$

$$\times\left\{1-\exp\left(-\frac{rv}{2\kappa}\left[\left(\frac{R+1}{2}\right)^{\frac{1}{2}}-1\right]\right)\right.$$

$$\left.\times\cos\left[\omega t-\frac{rv}{2\kappa}\left(\frac{R-1}{2}\right)^{\frac{1}{2}}\right]\right\}+T_a \tag{46}$$

where

$$R=(1+16\omega^2\kappa^2/v^4)^{\frac{1}{2}}.$$

For the author the chief interest in this result lies in the fact that it provides a formula for the phase velocity of the temperature wave set up by the periodic source. Thus, for a constant phase,

$$\omega t-\frac{rv}{2\kappa}\left(\frac{R-1}{2}\right)^{\frac{1}{2}}=\text{const.}$$

† The author has actually carried out the difficult calculations outlined in this program, before the simplicity of the result led him to the considerations that follow.

Therefore, the phase velocity, V, by differentiation of this equation, is

$$V=dr/dt=(2\omega\kappa/v)(2/R-1)^{\frac{1}{2}}. \quad (47)$$

For $v\to 0$,

$$V\to V_0=(2\omega\kappa)^{\frac{1}{2}}, \quad (48)$$

which is the phase velocity of the temperature wave when the fluid is at rest.

Frequently $v\gg V_0$. In that case it is convenient to expand the right-hand side of Eq. (47) in powers of V_0/v. We thus obtain:

$$V=v[1+\tfrac{1}{2}(V_0/v)^4-\tfrac{5}{8}(V_0/v)^8+\cdots]. \quad (49)$$

We see that under these circumstances it is practical to assume $V=v$; that is, that the temperature wave moves with the speed of the fluid. Even when the speed of the fluid is only five times V_0, so that the speed of the temperature wave in resting fluid is 20 percent of the speed of flow, the error in assuming

$$V=v$$

is less than 0.1 percent.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

On the Accuracy of the Long-Range Ballistic Rocket

W. E. Frye
The RAND Corporation, Santa Monica, California
(Received September 27, 1950)

The accuracy of a V-2 type ballistic rocket which is guided to the termination of propulsion and is in free flight thereafter is investigated. The effect of the rotation of the earth and the re-entry into the atmosphere on the accuracy is considered and the errors due to lack of thrust control are analyzed.

INTRODUCTION

BALLISTIC rockets designed for ranges greater than several hundred miles travel outside the atmosphere during most of the flight after the end of propulsion, and the trajectory during this time is uniquely determined by the position and velocity of the missile at the end of propulsion. A knowledge of the gravitational field of the earth over the path and the velocity of rotation of the earth then allows the calculation of the trajectory relative to the earth. As a good first approximation in considering the accuracy of such a rocket, the earth's gravitational field can be assumed to be central. In an inertial frame of reference the free-flight trajectory is then a planar ellipse with one focus at the earth's center. If the rotation of the earth is neglected (except for variations in the time of flight, its effect is systematic and calculable), the effect of variations in initial position and velocity on the trajectory can then be calculated in a straightforward manner.

When the missile re-enters the atmosphere, it is subject to drag forces due to the air whose effect can be approximated by integration of the equations of motion in a standard atmosphere. The missile will not in general follow this trajectory through the atmosphere because of the approximate nature of the equations of motion and because of nonstandard conditions such as winds that might be encountered in the atmosphere.

The method of error analysis used here assumes that the errors in the components of initial position and velocity are small. The effects of variations in each of the initial components on the range and deflection of the rocket trajectory are calculated separately. The individual errors are then suitably combined by adding variances.

EFFECT OF ERRORS IN INITIAL POSITION

Assume rectangular coordinate axes so oriented that the z direction is vertically upward at the initial point (end of propulsion), the y direction is horizontal in the plane of the trajectory, and the x direction is horizontal and normal to the trajectory plane.

Let Δy_0 = variation of initial position in the y direction, and $\Delta r(y_0)$ = resulting variation in range measured horizontally in the trajectory plane at the terminal point where the trajectory intersects the reference sphere (defined as the sphere with center at the center of the earth and passing through the initial point). Then, to the first-order approximation,

$$\Delta r(y_0)=\Delta y_0. \quad (1)$$

In terms of standard deviations, the range error at the terminal point due to an error in initial position in the y direction is then

$$\sigma_r(y_0)=\sigma(y_0). \quad (2)$$

The effect of a variation Δz_0 in initial position in the vertical z direction is to change the radius of the reference sphere by the amount Δz_0. At the terminal point on the reference sphere the missile velocity has the same magnitude v_0 as at the initial point (end of pro-

pulsion) and makes the same angle Ψ_0 with the vertical there as at the initial point. To the first order (for Δz_0 small) the resulting variation in range is

$$\Delta r(z_0)=\Delta z_0 \tan\Psi_0=\dot{y}_0\Delta z_0/\dot{z}_0, \tag{3}$$

where $\dot{y}_0$ and $\dot{z}_0$ are the y and z components of the velocity v_0 at the initial point. In terms of standard deviations the range error due to a vertical error in initial position is

$$\sigma_r(z_0)=\dot{y}_0\sigma(z_0)/\dot{z}_0. \tag{4}$$

To the first order a variation in initial position in the x direction results in a rotation of the trajectory plane about an axis parallel to the initial velocity vector and passing through the center of the earth, and the resulting variation at the terminal point is a deflection variation (in the direction normal to the trajectory plane) only and is given by

$$\Delta d(x_0)=\Delta x_0[\cos2\alpha-\cot\Psi_0 \sin2\alpha],$$

and the deflection error due to initial position error is

$$\sigma_d(x_0)=\sigma(x_0)[\cos2\alpha-\cot\Psi_0 \sin2\alpha], \tag{5}$$

where α is the semiangular range defined in the next section.

There is no contribution to range variation from Δx_0 and no deflection contributions from Δy_0 and Δz_0 to the first order.

The standard deviations of the errors in range and deflection at the terminal point due to errors in initial position are then

$$\sigma_r(p)=[\sigma(y_0)^2+\tan^2\Psi_0\cdot\sigma(z_0)^2]^{\frac{1}{2}} \tag{6}$$

assuming the errors in y_0 and z_0 are independent, and

$$\sigma_d(p)=[\cos2\alpha-\cot\Psi_0 \sin2\alpha]\sigma(x_0). \tag{7}$$

EFFECT OF ERRORS IN INITIAL VELOCITY

The semi-angular range α, defined as one-half the angle at the center of the earth between the initial and terminal points of the trajectory, can be expressed in terms of the y and z components of initial velocity and the magnitude of the escape velocity[1] v_e at the initial point by the equation,

$$\cot\alpha=(v_e^2/2\dot{y}_0\dot{z}_0)-\dot{y}_0/\dot{z}_0. \tag{8}$$

From Eq. (6) one obtains for the variation in range angle α due to variation of the y and z components of initial velocity

$$\Delta\alpha=A\Delta\dot{y}_0+B\Delta\dot{z}_0,$$

where

$$A=\left(\frac{v_e^2}{2\dot{y}_0^2\dot{z}_0}+\frac{1}{\dot{z}_0}\right)\sin^2\alpha,$$

$$B=\left(\frac{v_e^2}{2\dot{y}_0\dot{z}_0^2}-\frac{\dot{y}_0}{\dot{z}_0^2}\right)\sin^2\alpha.$$

[1] W. D. Macmillan, *Theoretical Mechanics, Statics and the Dynamics of a Particle* (McGraw-Hill Book Company, Inc., New York, 1927), p. 221.

The standard deviation of the error in α due to errors in initial velocity components is then

$$\sigma_\alpha=[A^2\sigma(\dot{y}_0)^2+B^2\sigma(\dot{z}_0)^2]^{\frac{1}{2}}, \tag{9}$$

assuming the errors in $\dot{y}_0$ and $\dot{z}_0$ to be independent. The resulting range error at the terminal point is then

$$\sigma_r(v_0)=2R\sigma_\alpha, \tag{10}$$

where R is the radius of the reference sphere.

A variation $\Delta\dot{x}_0$ in the initial lateral velocity results in a rotation of the trajectory plane and a change in the azimuthal direction θ of the initial velocity of amount

$$\Delta\theta=\Delta\dot{x}_0/\dot{y}_0.$$

The resulting deflection variation at the terminal point is

$$\Delta d(\dot{x}_0)=R \sin2\alpha\cdot\Delta\theta=R \sin2\alpha\Delta\dot{x}_0/\dot{y}_0.$$

Hence the deflection error due to an initial velocity error in the x direction is

$$\sigma_d(\dot{x}_0)=R \sin2\alpha\sigma(\dot{x}_0)/\dot{y}_0. \tag{11}$$

To the first order there is no contribution to the range variation from $\Delta\dot{x}_0$ and no contribution to the deflection variation from $\Delta\dot{y}_0$ and $\Delta\dot{z}_0$.

The total errors at the terminal point due to errors in initial position and velocity are then

$$\sigma_r=[\sigma(y_0)^2+4R^2A^2\sigma(\dot{y}_0)^2+(\dot{y}_0^2/\dot{z}_0^2)\sigma(z_0)^2+4R^2B^2\sigma(\dot{z}_0)^2]^{\frac{1}{2}}, \tag{12}$$

$$\sigma_d=[(\cos2\alpha-\cot\Psi_0 \sin2\alpha)^2\cdot\sigma(x_0)^2+(R^2 \sin^2 2\alpha/\dot{y}_0^2)\sigma(\dot{x}_0)^2]^{\frac{1}{2}}. \tag{13}$$

These errors in range and deflection are uncorrelated if the x, y, and z components of the errors in initial position and velocity are independent.

EFFECT ON ERRORS OF THE ROTATION OF THE EARTH

The rotation of the earth introduces a dependence on the time-of-flight into the errors at the terminal point on the reference sphere. If the variation of the time-of-flight with the variations (errors) in initial position and velocity is known, then the contribution to the error at the terminal point due to the rotation of the earth is obtained by multiplying the velocity of the end point (due to the earth's rotation) by the variation in the time-of-flight.

To the first order, variation in initial position in the x and y (horizontal) directions will not change the time of flight. The variation in time of flight t_f due to a variation in the z (vertical) direction is given approximately by the quotient of the variation of the reference sphere radius and the vertical component of the rocket velocity at the terminal point or $\Delta z_0/\dot{z}_0$. Variation in $\dot{x}_0$ has no effect on t_f to the first order. The effect of variations in $\dot{y}_0$ and $\dot{z}_0$ on t_f can be obtained by consideration of the equation for the elapsed time on an

elliptic orbit.[2] The errors due to the variation in time of flight are in general small compared to the other errors.

THE ERRORS AT IMPACT

The errors at the terminal point on the reference sphere can be translated down to impact on the surface of the earth by consideration of the equation of motion[3]

$$(d^2\mathbf{r}/dt^2)+gR^2\mathbf{r}/r^3=\mathbf{b}(t),$$

where $\mathbf{r}$ is the position vector of the missile's center of mass with respect to the center of the earth, $\mathbf{b}(t)$ is the acceleration of the missile's center of mass due to nongravitational forces, and g is the acceleration of gravity at the distance R from the center of the earth. Using the linearized approximation[4] to the gravitational force, one obtains the differential equations for the variations Δd and Δr in the deflection and range directions as the missile travels through the atmosphere from the terminal point to impact:

$$\Delta d''+k^2\Delta d=\beta_d(t), \tag{14}$$

$$\Delta r''+k^2\Delta r=\beta_r(t), \tag{15}$$

where $k^2=g/R$ and $\beta_d(t)$ and $\beta_r(t)$ are the components of the perturbations in $\mathbf{b}(t)$ caused by nonstandard forces such as winds in the atmosphere. The ballistics of the missile in a "standard" atmosphere must of course be determined in advance in order to calculate the trajectory in such an atmosphere.

Assuming that the $\beta_{d(r)}(t)$ are random variables and utilizing the symmetry in velocity at the initial and terminal points, one obtains from the well-known solutions of Eqs. (14) and (15) the errors at impact at the surface of the earth:

$$\sigma_{d_e}=[\cos^2kt_1\cdot\sigma_d^2+k^{-2}\sin^2kt_1\cdot\sigma(\dot{x}_0)^2+k^2\langle P_d^2\rangle]^{\frac{1}{2}}, \tag{16}$$

$$\sigma_{r_e}=[\cos^2kt_1\cdot\sigma_r^2+k^{-2}\sin^2kt_1\cdot\sigma(\dot{y}_0)^2+k^2\langle P_r^2\rangle]^{\frac{1}{2}}, \tag{17}$$

where

$$P_{d(r)}=\int_0^{t_1}\beta_{d(r)}(\tau)\sin k(t_1-\tau)d\tau,$$

and the angular brackets denote expected values with respect to the probability distribution of the $\beta_{d(r)}(t)$. It is assumed that the mean values of the position and velocity errors at the terminal point and of the $\beta_{d(r)}(t)$ are zero, while t_1 is the time of flight from the terminal point to impact on the surface of the earth. An approximate value for t_1 is given by $h/\dot{z}_0$, where h is the altitude of the terminal point above the earth.

ERROR DUE TO LACK OF THRUST CONTROL

If the rocket's trajectory during the propulsion phase is programmed without control of thrust, as in the case of the V-2 rocket, and if the propulsion is terminated when a pre-assigned value of the velocity is attained, errors in the initial position of the rocket will arise because of variation of the thrust from the standard design value during propulsion. Variation in thrust can occur because of variation in the fuel flow rate and variations in the specific impulse. If f is the fuel flow rate, then the change in path length ΔS from launching to end of propulsion caused by a constant change in the fuel flow rate is given approximately by[5]

$$\Delta S=-S\Delta f/f.$$

The change in path length due to a constant change ΔI in specific impulse is given approximately by[5]

$$\Delta S=(S-v_0^2/a_0)\Delta I/I,$$

where a_0 = acceleration at end of propulsion, v_0 = velocity at end of propulsion, I = specific impulse.

For any particular rocket launching the fuel flow rate and the specific impulse can be considered approximately constant, and the Δf and ΔI will be the deviations from the design values which are used in trajectory calculations. The standard deviation of S due to variations in f and I among many rockets is

$$\sigma_S=[S^2(\sigma_f^2/f^2)+(S-v_0^2/a_0)^2\sigma_I^2/I^2]^{\frac{1}{2}}, \tag{18}$$

since the variations in f and I are independent. In using this equation, one can approximate S by the slant range from the launching point to the initial point.

The change in path length ΔS results in a change in the initial position at end of propulsion given by

$$\Delta y_0=\Delta S\sin\Psi_0,$$
$$\Delta z_0=\Delta S\cos\Psi_0,$$

in the horizontal and vertical directions, respectively. This results, by Eqs. (1) and (3), in a variation in range at the terminal point of

$$\Delta r=2\sin\Psi_0\Delta S.$$

In terms of standard deviation of the resulting error

$$\sigma_r=2\sin\Psi_0\cdot\sigma_S. \tag{19}$$

There will be no first-order contribution to the deflection error at the terminal point due to lack of control of thrust.

This contribution to the range error at the terminal point due to lack of thrust control could be reduced significantly if control of the rocket motor thrust is possible during the propulsion phase, since then the trajectory and velocity of the missile could be programmed in space limited only by the accuracy of the control and the accuracy of position and velocity measurements.

APPLICATION TO THE V-2 ROCKET

The results of the previous sections can be applied to the German V-2 rocket. Descriptions of this ballistic missile have been widely published,[6] and sufficient in-

[2] Reference 1, pp. 280, 285.
[3] Gilvarry, Browne, and Williams, J. Appl. Phys. **21**, 753 (1950).
[4] Reference 3, p. 756.
[5] See Appendix I.
[6] For example, J. M. J. Kooy, and J. W. H. Uytenbogaart, *Ballistics of the Future* (McGraw-Hill Book Company, Inc., New York, 1946); W. G. A. Perring, J. Roy. Aeronaut. Soc. **50**, 483 (1946); T. M. Moore, Elec. Eng. **68**, 996 (1949).

TABLE I.

n	Cause of error	$\sigma_n(x_0)$ m	$\sigma_n(y_0)$ m	$\sigma_n(z_0)$ m	$\sigma_n(\dot{x}_0)$ m/sec	$\sigma_n(\dot{y}_0)$ m/sec	$\sigma_n(\dot{z}_0)$ m/sec
1	Alignment	125	···	97	5.52	···	4.64
2	Gyro drift	233	···	233	13.82	···	11.62
3	Integrating accelerometer error	···	···	···	···	12.26	10.29
4	Uncertainty in fuel cut-off time	···	613	514	···	11.26	9.45
5	Deviation in specific impulse and fuel flow rate	···	4295	3606	···	···	···
	$(\Sigma_n\sigma_n^2)^{\frac{1}{2}}=$	264	4338	3651	14.88	16.65	18.75

formation is available and reasonable assumptions can be made to permit an estimate of its accuracy. The following data on the V-2 trajectory are required for the calculations: Range = 300 km (firings on London); azimuth, $\theta \cong 270°$; speed at fuel cutoff, $v_0 = 1600$ m/sec; angle of velocity with vertical at fuel cutoff, $\Psi_0 = 50°$; altitude at fuel cutoff, $h = 27$ km; range at fuel cutoff = 20 km; duration of propulsion, $t_0 = 60$ sec; acceleration at fuel cutoff, $a_0 = 3g$ (reduced thrust program); and time-of-flight after fuel cutoff, $t_f = 240$ sec.

The guidance and control system that was used operationally in the V-2 consisted of two gyroscopes for stabilization including a pitch programming mechanism, an integrating accelerometer system for determining the fuel cut-off point by integration of the measured acceleration to obtain the velocity, and an electromechanical-hydraulic servo system for controlling the missile. A radio system for controlling the lateral motion was developed for the V-2 but was apparently not used in the firings on London. A gyro-stabilized platform on which a lateral integrating accelerometer in addition to a range accelerometer were mounted was developed but probably did not attain operational use.

It is necessary to make certain assumptions about the accuracy of the guidance components in the V-2 in order to calculate the errors in position and velocity at the fuel cut-off point. The most important contributions to the errors are due to the initial misalignment of the stabilization gyros, the drift of the gyros with time, the inaccuracy of the integrating accelerometer system, the uncertainty in the time of fuel cutoff, and the deviation of specific thrust and fuel flow rate from standard. Each of these independent sources of error will be discussed and its effect on the errors at the end of propulsion evaluated.

1. Initial Misalignment of Gyros

The "pitch" gyro is oriented with its rotation axis vertical, and during propulsion is precessed according to the predetermined pitch program. The other gyro controls the missile in roll and yaw, and its rotation axis is oriented horizontally and normal to the target direction. The probable error in aligning the gyros was claimed to be 5 minutes of arc. Accordingly, it will be assumed that the standard deviation of the alignment errors is 10 minutes of arc. Assuming that the alignment errors for the pitch, yaw and roll axes are independent, one obtains for the approximate first-order position errors at the fuel cut-off point due to the alignment errors:

$$\sigma_1(x_0) = (r_s^2 + h^2)^{\frac{1}{2}}\sigma_A,$$
$$\sigma_1(z_0) = r_s\sigma_A,$$

where r_s is the slant range from the launching point to the cut-off point, h is the altitude of the cut-off point, and σ_A is the standard deviation of the alignment error.

The approximate first-order velocity errors at the cut-off point due to the alignment errors are

$$\sigma_1(\dot{x}_0) = (v_0^2 + \dot{z}_0^2)^{\frac{1}{2}}\sigma_A,$$
$$\sigma_1(\dot{z}_0) = v_0\sigma_A.$$

2. Drift of Gyros

The magnitude and direction of drift of the gyros will vary from missile to missile. The distribution of drift rates for many missiles will be assumed to have a mean value of zero and a standard deviation σ_D of 25 degrees per hour. The latter figure is an estimate based on comparable American units with allowance for the high acceleration encountered in the V-2.

An approximate value for the position error due to gyro drift can be obtained by a simple calculation. If S represents the distance along the trajectory, then for the error $\sigma_2(x)$ in the x direction

$$d\sigma_2(x)/dS = \tan(\sigma_D t) \simeq \sigma_D t.$$

Approximate the velocity along the trajectory by a linear function

$$dS/dt = at,$$

where

$$a = v_0/t_0.$$

Then the error at the end of propulsion can be written

$$\sigma_2(x_0) = \int_0^{t_0} \frac{d\sigma_2(x)}{dS}\frac{dS}{dt}dt = \frac{\sigma_D v_0 t_0^2}{3}.$$

An identical expression occurs for the error in z_0.

The approximate velocity errors at cutoff due to the gyro drifts are

$$\sigma_2(\dot{x}_0) = \sigma_D t_0 (v_0^2 + \dot{z}_0^2)^{\frac{1}{2}},$$
$$\sigma_2(\dot{z}_0) = \sigma_D t_0 v_0.$$

3. Error in Integrating Accelerometer System

The error in measuring the velocity by means of the integrating accelerometer system is claimed to be the order of $\frac{1}{2}$ to 1 percent from both laboratory and flight tests. If the standard deviation of the error is assumed to be 1 percent of the velocity, then the velocity errors at cutoff caused by inaccuracy of the integrating accelerometer system will be

$$\sigma_3(\dot{y}_0) = 0.01 v_0 \sin\Psi_0,$$
$$\sigma_3(\dot{z}_0) = 0.01 v_0 \cos\Psi_0.$$

4. Uncertainty in Time of Fuel Cutoff

Approximately two seconds were required for cutting off the fuel flow and reducing the thrust to zero.[7] As-

[7] W. G. A. Perring, reference 6, p. 501.

suming a standard deviation of 0.5 sec as the uncertainty in the time of cessation of thrust, σ_t, one obtains for the errors in velocity at the end of propulsion

$$\sigma_4(\dot{y}_0)=a_0\sigma_t\sin\Psi_0,$$
$$\sigma_4(\dot{z}_0)=a_0\sigma_t\cos\Psi_0.$$

There will also be errors in position due to σ_t as follows:

$$\sigma_4(y_0)=v_0\sigma_t\sin\Psi_0,$$
$$\sigma_4(z_0)=v_0\sigma_t\cos\Psi_0.$$

5. Deviation in Specific Impulse and Fuel Flow Rate

A possible variation of 10 percent in the specific impulse of the V-2 rocket motor has been claimed. If one assumes a standard deviation of 10 percent of the specific impulse and a standard deviation of 5 percent of the fuel flow rate, then the errors in position at the termination of propulsion due to the deviations in specific impulse and fuel flow rate will be

$$\sigma_5(y_0)=\{[0.05S]^2+[0.1(S-v_0^2/a_0)]^2\}^{\frac{1}{2}}\sin\Psi_0,$$
$$\sigma_5(z_0)=\{[0.05S]^2+[0.1(S-v_0^2/a_0)]^2\}^{\frac{1}{2}}\cos\Psi_0,$$

from the equations derived in Appendix I, where S can be approximated by the slant range from the launching point to the point of fuel cutoff.

The numerical errors in initial position and velocity caused by these various causes are shown in Table I. Using Eqs. (12) and (13) one finds for the range and deflection errors at the terminal point due to errors in initial position and velocity

$$\sigma_r=8.7\text{ km},$$
$$\sigma_d=3.6\text{ km}.$$

The circular probable error is then approximately

$$\text{CEP}=(\tfrac{1}{2}\log_e 2)^{\frac{1}{2}}(\sigma_r+\sigma_d)=7.2\text{ km}.$$

If the atmospheric perturbations $\beta(t)$ are neglected in Eqs. (16) and (17), the errors at the target are essentially the same as the above figures for σ_r and σ_d.

A statistical analysis of the impacts of V-2s in the London area during the war (with impacts having gross errors apparently due to malfunctioning eliminated) showed the following standard deviations:

$$\sigma_r=13\text{ km},$$
$$\sigma_d=11\text{ km}.$$

The discrepancy in magnitudes can perhaps be attributed to perturbations in the atmosphere which were ignored in the calculations and to an inevitable degradation of performance of a new weapon under military operational conditions. The possibility of deliberate dispersion or shift in the aiming point by the Germans cannot be eliminated from consideration.

The later firings of the V-2 on Brussels at a range of 200 km showed improved accuracy in both range and deflection, the latter being attributed to the use of a ground-based radio beam system to control the lateral motion during propulsion.

APPENDIX I

EFFECT OF THRUST PERTURBATIONS DURING PROPULSION[8]

The thrust of a rocket propulsion system is proportional to the product of the specific impulse I and the fuel flow rate f.[9] The acceleration of the missile is then proportional to

$$If/1-ft,$$

where t is the time from launching and the denominator represents the mass of the missile as a function of time. The thrust is, on the average, the largest force acting on the missile. Hence, the path length S is given approximately by the double integration of the acceleration

$$S\propto(I/f)[(1-ft)\log(1-ft)+ft],$$

assuming I and f are essentially constant during propulsion. Similarly, the velocity v is given by a single integration of the acceleration,

$$v\propto -I\log(1-ft).$$

If the velocity at the end of propulsion is always constant, then the product $f\cdot t_0$ is constant, where t_0 is the time to the end of propulsion. Hence, at the end of propulsion

$$S(t_0)\cdot f=\text{constant};$$

and for a change Δf in fuel-flow rate during propulsion, the resulting change in path length ΔS is given by

$$S\cdot\Delta f+f\cdot\Delta S=0$$

or

$$\Delta S=-S\Delta f/f.$$

Since the acceleration is proportional to

$$If/1-ft,$$

the effect of a change $\Delta I/I$ in the specific impulse is given by

$$\Delta I/I=\Delta a/a,$$

where a is the acceleration of the missile. The perturbed acceleration due to a constant percentage change $\Delta I/I$ is

$$a'=a+\Delta a=(1+\Delta I/I)a.$$

The perturbed path length S' is given by a double integration of the perturbed acceleration

$$S'=\int_0^t\int_0^t a'dt^2=(1+\Delta I/I)\int_0^t\int_0^t adt^2=(1+\Delta I/I)S.$$

Similarly, the perturbed velocity is

$$v'=\int_0^t a'dt=(1+\Delta I/I)v.$$

The perturbed path length to the same velocity v_0 is $S'(t_0)$ decreased by $v_0\cdot\Delta t$ (assuming $\Delta I/I$ is positive), where

$$\Delta t=\Delta v_0/a_0$$

is the perturbation in time to reach the velocity v_0 at the end of propulsion, a_0 is the acceleration at the end of propulsion, and Δv_0 is given by

$$\Delta v_0=v_0'-v_0=\Delta I v_0/I.$$

Hence the change in path length is

$$\Delta S=(S'-v_0\cdot\Delta t)-S$$

or

$$\Delta S=(S-v_0^2/a_0)\Delta I/I$$

owing to the change $\Delta I/I$ in specific impulse.

[8] The results contained in this appendix are due to Mr. K. I. Martin of The RAND Corporation.

[9] J. M. J. Kooy and J. W. H. Uytenbogaart, reference 6, p. 396.

Probe Technique for the Measurement of Electron Temperature

M. A. EASLEY
General Electric Company, Lamp Development Laboratory, Nela Park, Cleveland, Ohio
(Received October 26, 1950)

Discontinuities in the slopes of Langmuir probe characteristics for discharges in mercury vapor in the presence of argon or krypton gas were eliminated by heating the wire probe by electron bombardment before each reading. The results indicate that in many cases nonlinear probe plots are due to change in the contact potential of the probe and do not mean a non-maxwellian distribution of electron velocities. Typical probe plots are given for discharges with and without striations, and also for a discharge in gas contaminated by a trace of CO or CO_2. Electron temperature data are listed for representative pressures of gas and mercury vapor and for various discharge currents.

INTRODUCTION

THE electrical conditions in the positive column of a discharge are specified by the positive ion current, the electron concentration, and the electron energy distribution. The usual method of studying these characteristics makes use of probes placed in the arc stream. Langmuir[1] has showed that the logarithm of the electron current to a probe negative with respect to space is a linear function of the probe voltage provided the electrons in the plasma have a maxwellian velocity distribution. Thus if the plot of the logarithm of probe current (i_p) *versus* probe voltage (V_p) is a straight line for V_p negative with respect to space, the electron energies are distributed according to the Maxwell law and can be described by an electron temperature T_e obtained from the slope (e/kT_e) of the straight line. Experimental data, however, often have given plots that were nonlinear, or that consisted of two or more straight lines. The question then arises as to whether the nonlinearity is due to a non-maxwellian distribution of electron velocities, to the superposition of two or more maxwellian distributions, or to flaws in probe technique. The object of the work reported in this paper was to investigate the causes for nonlinear probe characteristics obtained for certain mercury discharges.

APPARATUS

The experiments were carried out with cylindrical Pyrex discharge tubes having diameters from 34 to 50 mm. The cathodes were oxide-coated tungsten filaments of the type used in G.E. sodium vapor lamps and were run at a current of about 6.5 amp. The anodes were nickel cylinders, having a diameter about 75 percent that of the Pyrex tube and a length of 15 to 25 mm. The discharge tubes each had two cylindrical probes of tungsten wire, either 5-mil diameter and 10-mm exposed length, or 3-mil diameter and 5-mm length. These wire probes were centered in the discharge with their axis perpendicular to the tube axis. Opposite each wire probe was a wall probe consisting of a molybdenum disk 4.85 mm in diameter, surrounded by a guard ring, with mica insulation between the metal and the Pyrex wall. The probe dimensions were such that the maximum electron current to the probes was never more than about 10 percent of the discharge current.

In preparation for measurements, the tube was baked at an oven temperature of 520°C. The anode and wire probes were cleaned with high frequency heating. The cathode was degassed by running it two to three hours at currents of from 6 to 8 amp, with the gas evolved being pumped off as quickly as possible. This treatment shortened the life of the cathode, but was necessary for the elimination of objectionable traces of CO in the finished lamp. Mercury was distilled into the tube, which contained a few millimeters of argon or krypton gas. The wall probes then were cleaned by running a discharge to them, using, in turn, each disk and each guard ring as anode, with the discharge current adjusted to raise the temperature of the metal to about 1200°C. After the tube was filled with argon or krypton gas and arced several times, it was pumped out thoroughly, filled to the desired pressure with spectroscopically pure gas, and tipped off the vacuum system. It was necessary then to run the discharge at 400 ma for 5 to 10 hours to clean up remaining impurities.

For accurate control of mercury pressure, the tube was operated in a well-stirred water bath with temperature thermostatically controlled within ±0.2°. The usual technique of taking probe measurements was employed, with the probe voltage set at a desired value with respect to either the anode or the cathode, and the corresponding probe current recorded. In general, the anode was used for reference, but for discharges with anode disturbances, the cathode was used. Since the cathode was large, it was not affected by the probe current and was satisfactory as the reference electrode until it was nearly deactivated. An oscilloscope was connected between the electrode and probe to detect the presence of oscillations. Except where otherwise indicated, the measurements recorded represent discharges in which oscillation amplitudes at frequencies below 1 mc were less than 0.04 v.

EXPERIMENTAL RESULTS

When the relation between probe current and voltage was obtained by gradually making V_p more negative

[1] Langmuir, Gen. Elec. Rev. **26**, 731 (1923).

with respect to the space potential V_s, the $\log i_p - V_p$ characteristics for discharges in mercury vapor in argon or krypton gas often consisted of a series of lines. The first discontinuity in the plot usually occurred with the probe 3 or 4 volts negative with respect to V_s. The slope of the line near V_s was always less than that of the second line, with the difference in slope greater at low discharge currents than at high. It was noted that for some discharges, the value of space potential (i.e., the voltage at which the relation $\log i_p = kV_p$ fails) depended upon the method used in taking the data. When the run was started with the probe positive with respect to space, and probe current was recorded as probe voltage was gradually made more negative, the space potential indicated by the plot was different from that for a run taken with probe voltage gradually changed from -50 v to V_s. This shift in the potential of the plasma with respect to the reference electrode suggested the possibility of a change in the contact potential between probe and electrode.

In an attempt to maintain a constant mercury-tungsten surface, wire probes were bombarded by positive ions before each $i_p - V_p$ reading.[2] After such bombardment, electron current to the probe decreased too rapidly for accurate measurement. When the tungsten probes were heated to incandescence by electron bombardment before each reading, the current drawn by the probe was constant for a short interval and could be measured. Figure 1 compares a probe characteristic obtained in this way with characteristics obtained by gradually changing the probe voltage. Runs *A–B–C–D* were taken as rapidly as possible, with no pause between them. When probe voltage gradually was made more negative with respect to space (Fig. 1*A*), the characteristic showed a change in slope at -3.8 v from space, with the slope near space potential corresponding to an electron temperature of 12,100°K. The characteristic obtained as probe voltage was decreased from -60 v had two breaks, at about -1 and -3.9 v from space (Fig. 1*B*). With the probe heated to incandescence by electron bombardment before each reading (Fig. 1*C*), the characteristic was linear over the measurable range of electron current to the probe, or for 9.5 v from space potential, and had a slope corresponding to an electron temperature of 11,500°K. Run *D*, a repeat of Run *A*, was started immediately after *C*, with the probe surface still clean from the treatment in Run *C*. The data from Run *D* lie on a smooth curve from V_s (-13.2 v) to -19 v, and cannot be represented by straight lines.

The results shown in Fig. 1 are typical of many runs taken on well-seasoned tubes containing argon or krypton gas with various pressures of mercury vapor, and operating at currents from 200 to 1500 ma. At the lowest currents, the probe voltage had to be several volts positive with respect to space potential in order to have sufficient electron bombardment of the probe.

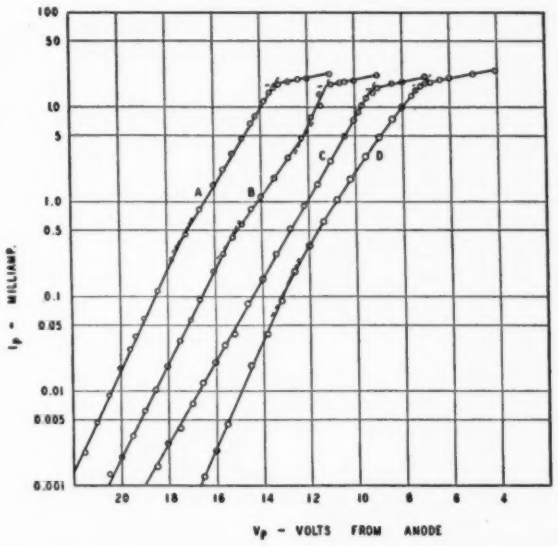


FIG. 1. Probe characteristics for 400 ma dc discharge. Tube 34 mm inside diameter, with 3.5 mm argon and 6.1μ-mercury (40° bath). Wire probe (5 mm×3 mil tungsten). Curve *A*: Probe voltage changed gradually from -9 to -60 v with respect to the anode. Curve *B*: Probe voltage changed from -60 to -9 v with respect to the anode. Curve *C*: Probe heated by electron bombardment before each reading. Curve *D*: Repeat of Curve *A*, taken immediately after Curve *C*. Curves *B–C–D* are each displaced 2 v with respect to the preceding curve.

Thus the disturbance of the arc stream by the probe current during the heating cycle was greater than at higher currents, and the results were less accurate. With the wire probes used in these experiments, the maximum current drawn by the probes during the heating cycle was less than 7 percent of the discharge current for the arc operated at 400 to 1500 ma.

The electron current to the wall probes was never sufficient to heat them to incandescence. The probe characteristics, therefore, were similar to those shown in plots *A* and *B*, Fig. 1. Tests made to determine the effect of sputtering the wall probes likewise resulted in nonlinear plots.

Since results similar to those shown in Fig. 1 were obtained with several different tubes containing argon or krypton gas and mercury vapor at different pressures, it is logical to attribute these characteristics to mercury rather than to the filling gas or to impurities. Kenty[3] has described a "clean-up" of mercury by an iron cylinder, ascribing it to the penetration of the metal surface by mercury ions. If a similar mercury deposit were laid down over the tungsten wire, then the shift in space potential and the nonlinearity of the probe characteristics found in these experiments could be explained on the basis of a change in the contact potential of the probe with respect to the anode. Thus as the probe voltage became more and more negative, the probe surface would change from tungsten to tungsten-mercury. Electron bombardment of the probe would

[2] C. S. Cummings and L. Tonks, Phys. Rev. **59**, 514 (1941).

[3] C. Kenty, J. Appl. Phys. **9**, 765 (1938).

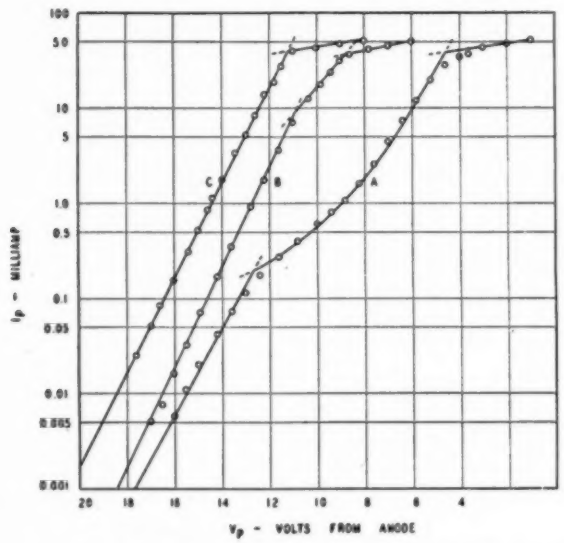


FIG. 2. Probe characteristics for 400 ma dc discharge. Tube 34 mm inside diameter with 1.85 mm krypton, 6.1μ mercury (40° bath). Wire probe (10 mm×5 mil tungsten) heated by electron bombardment before each reading. Curve *A*: Discharge with traveling striations. Curve *B*: Discharge in gas containing small amount of impurity, probably CO—CO_2. Curve *C*: Discharge in well-seasoned tube. Curves displaced with respect to each other along voltage axis.

drive off the mercury, and, when sufficiently intense, yield a clean tungsten surface. The difference between the space potential found with the probe voltage going from positive to negative and that with the probe voltage changing from negative to positive indicates that the contact potential of the mercury-tungsten surface was a few tenths of a volt more positive than that for tungsten. The fact that the slope of the characteristic in Fig. 1*C* lies between that of the upper and lower lines in plots 1*A* and 1*B* indicates some shift in contact potential in the upper as well as in the lower part of the plot. This conclusion is given support by Fig. 1*D*.

Nonlinear $\log i_p - V_p$ plots have been obtained for certain discharges, regardless of the technique used. Oleson and Found[4] have reported upon results for a discharge with traveling striations. Their results have been checked, using the bombarded probe technique. A typical curve is shown in Fig. 2*A*. The probe was made 2 v positive before each reading. Neither the curvature nor the break at $V_p = 18$ v was affected by this technique. The plot can be resolved into two maxwellian distributions, indicating two electron groups with electron temperatures of about 11,000° and 26,500°. The gradual break in the plot at V_s is typical for discharges with oscillations.

Discontinuities in the $\log i_p - V_p$ plot were found also for discharges containing traces of CO, CO_2, or C_2. It is estimated that the lamp represented by Curve *B*, Fig. 2, contained less than 0.5μ of CO or CO_2. The slope of the line obtained with probe voltage greater than −15.9 v was nearly the same as that of the probe characteristic for the discharge after thorough seasoning (Fig. 2*C*). Near space potential, however, the slope corresponded to an electron temperature about 5000°K higher than that for the discharge with no detectable contamination. Similar probe characteristics obtained with tubes on a vacuum system without a mercury cutoff between the lamp and the stopcock indicated that the small amount of hydrocarbons from the stopcock grease had a similar effect. It is probable that the nonlinearity of these plots is not indicative of a nonmaxwellian energy distribution, but is the result of a change in the probe contact potential similar to that caused by the mercury. As Fig. 2*B* shows, bombarding the probe before each measurement did not eliminate the discontinuity in the plot. The diffusion rate of the CO or CO_2 is rapid enough, however, that the surface of the probe could be changed before the voltmeter-ammeter readings were obtained.

DISCUSSION

A large number of discharges in mercury vapor mixed with pure argon or krypton gas have been studied, using the probe technique just described. Gas pressure was varied from 1 to 3.5 mm, mercury pressure from 0.5 to 46μ, and discharge current from 200 to 1500 ma. For discharges without oscillations or traveling striations, linear probe characteristics were obtained over the measurable range of probe currents with the wire probes, but not with the disk probes. Nonlinear characteristics were obtained for tubes containing small amounts of impurities. Typical data for well-seasoned tubes are given in Tables I and II, showing the variation of electron temperature with mercury pressure and with discharge current. With the usual method of gradually changing the probe voltage, nonlinear probe plots had been obtained for similar tubes. It is interesting to note that it was found possible in many cases to suppress

TABLE I. Electron temperature (°K) for a 400 ma dc discharge in mercury vapor mixed with argon or krypton gas.

Gas	Pressure (mm at 25°C)	Mercury pressure (Microns) 1.5	2.8	6.1	12.7	25
A	3.5	15,200°	13,700°	11,500°	10,000°	9800°
Kr	2	11,700°	11,700°	10,400°	8900°	8000°

TABLE II. Electron temperature (°K) *vs* discharge current for a dc discharge in 6.1μ-mercury vapor mixed with argon or krypton gas.

Current (milliamp)	Electron temperature (°K) for 3.5 mm A	2 mm Kr
200	12,200°	10,700°
400	11,500°	10,400°
800	10,500°	9400°
1500	10,000°	8800°

[4] N. L. Oleson and C. G. Found, J. Appl. Phys. **20**, 417 (1949).

oscillations and striations by bringing a small horseshoe magnet near the anode end of the tube. The use of the magnet extended the range of operation without oscillations to lower currents and lower mercury pressures.

Data similar to that given in the tables lead to the conclusion that in the discharges studied, the electron velocity distribution followed the maxwellian distribution law over the measurable range of electron current (i.e., to 9 to 11 volts from space potential).

ACKNOWLEDGMENT

The writer wishes to express appreciation to Dr. C. G. Found, in charge of research at Nela Park, for his encouragement and suggestions, to her colleagues Dr. B. T. Barnes and Dr. C. Kenty for many stimulating discussions, and, especially, to a former colleague, Dr. N. L. Oleson, for generously sharing his knowledge of positive column research and his experience in the use of probes.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Some Mechanical Properties of Graphite at Elevated Temperatures*

C. MALMSTROM, R. KEEN, AND L. GREEN, JR.
North American Aviation, Inc., Atomic Energy Research Department, Downey, California

(Received November 10, 1950)

The short-time tensile breaking strength of various grades of graphite was measured as a function of temperature from room temperature to the sublimation point. A characteristic common to all the strength *versus* temperature curves is that the strength approximately doubles in going from room temperature to about 2500°C and then decreases rapidly to zero near the sublimation point. Graphite can be used as a structural material at temperatures much higher than those permissible with other materials. The density distribution within an original block is given for grade EBP graphite. The creep characteristics of grade ECA graphite under tensile stress at elevated temperatures were investigated. The testing temperature ranged from 2200 to 2900°C while the applied stress varied from 500 to 5500 psi. The range of "steady" creep rates measured extended from 3×10^{-8} to 2×10^{-4} in. per in. per sec. The activation energy and change in entropy defined by rate theory were calculated for the above conditions. Low frequency dynamic measurements of the Young's modulus of grade ECA graphite made at temperatures from 1000 to 2000°C showed an increase of the modulus with temperature.

I. INTRODUCTION

IN view of the interest in refractory materials for many high temperature applications, information on the properties of graphite up to the sublimation point was desired. A search of the literature and inquiry[1] with the National Carbon Company indicated that no studies on the mechanical properties of graphite had been made at temperatures above 1500°C, so an experimental investigation was undertaken. All the grades of graphite used in this investigation were manufactured by the National Carbon Company and are listed in Table I.[2]

The anisotropic nature of graphite is well known. The tensile breaking strength of extruded graphite, for example, is higher in the direction of extrusion than perpendicular to that direction. In molded graphite the strength is greatest in a direction normal to the molding pressure vector. Such anisotropy of the polycrystalline material is attributed to a preferred orientation assumed by the individual graphite laminas during the manufacturing process. All the specimens used in the recent investigation were fabricated with their axes parallel to the direction of maximum strength.

As the investigation progressed it became evident that many factors contributed to the observed variations in the properties of graphite. In the present preliminary survey, however, no attempt was made to investigate fully the effect of such variables as grain orientation, density, particle size, types of pitch and coke, and graphitization temperature. It must be emphasized, however, that these variables (which may vary from piece to piece within a given batch as well as from grade to grade) may profoundly influence the mechanical properties of the material.

TABLE I.[a]

Graphite grade	Method of manufacture	Graphitization temperature °C	Approximate percentage of oriented grains in final product	Final density g/cm³
AGX	Extruded	2600	40	1.58
C-18	Molded	2600	30	1.60
SA-25	Molded	3000	1	1.55
ECA[b]	Extruded	3000	60	1.67
EBP[c]	Molded	3000	30	1.76

[a] The densities as given by the National Carbon Company and presented in Table I can only be considered as average values. For example, it will be shown later that the density of EBP varies from 1.59 g/cm³ to 1.80 g/cm³ within one block.
[b] Now designated as AUF by National Carbon Company.
[c] Now designated as AWG by National Carbon Company.

*This paper is based upon studies conducted for the AEC under Contract AT-11-1-GEN-8.

[1] Private communication from H. L. Larson, National Carbon Company, Inc., Cleveland, Ohio, November, 1946 to September, 1947.

[2] Private communication from H. G. MacPherson, National Carbon Company, Inc., Cleveland, Ohio, October, 1948.

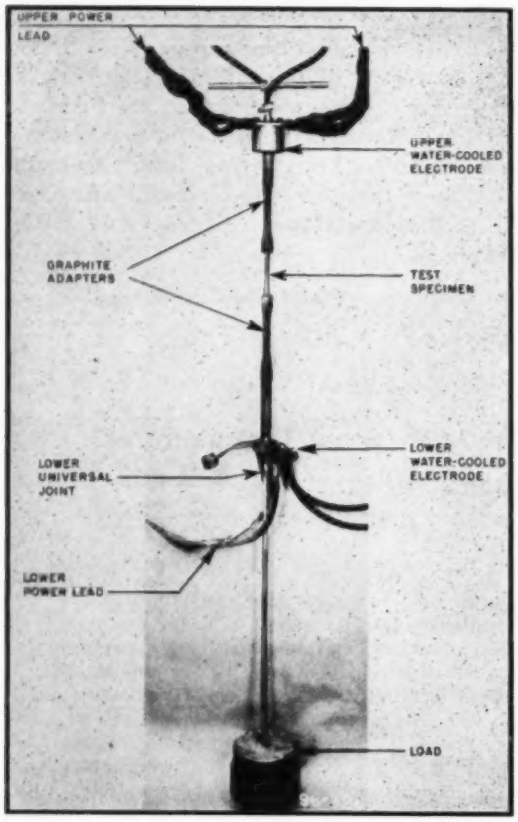


FIG. 1. Specimen, adapters, and electrodes used in high temperature furnace.

II. EXPERIMENTAL PROCEDURE

A. Tensile Breaking Strength

The tensile specimens used in the present investigation were heated by the passage of a direct current. The electrodes and graphite adapters supporting the specimen are shown in Fig. 1. This figure shows that the load was applied by weights placed in a pan. Later, the pan

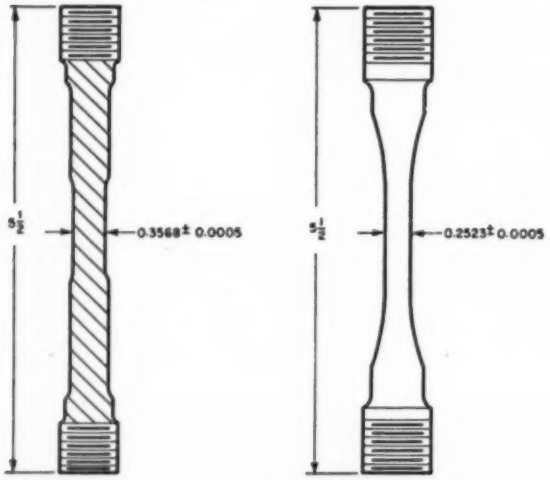


FIG. 2. Specimens utilized in strength measurements.

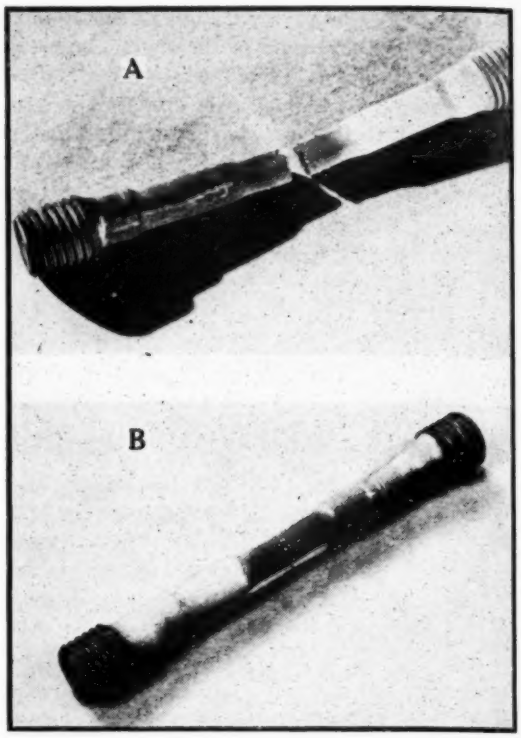


FIG. 3. Broken and unbroken specimens.

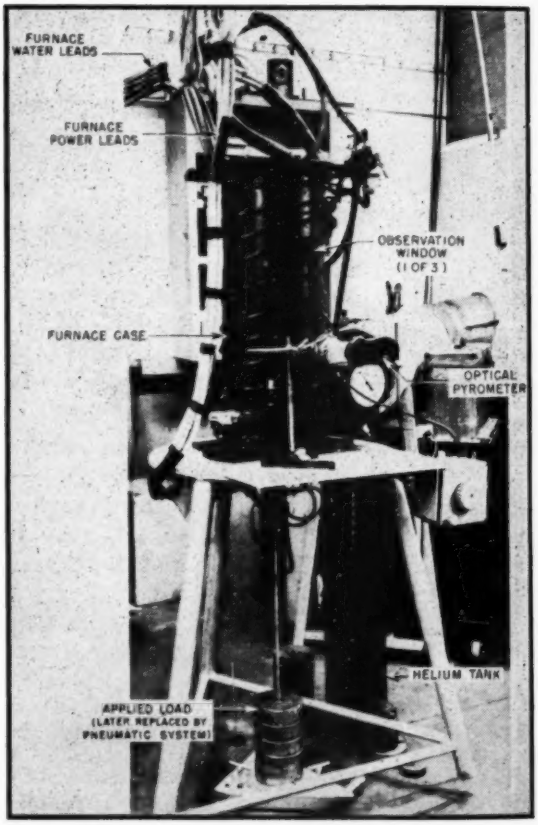


FIG. 4. High temperature furnace.

was replaced by a bellows arrangement and the load was applied pneumatically. Another method of application utilized a lever mechanism actuated by a pail of lead shot. Ball-bearing universal joints at the two electrodes insured the application of a pure tensile stress. Pertinent details of the two types of specimen which were utilized are given in Fig. 2. Typical specimens which had attained a very high temperature may be seen in Fig. 3. The beads on the surface apparently were formed by the condensation of material which had migrated through the porous structure after subliming at the center of the specimen.

The specimens were tested in a helium atmosphere contained by a furnace in which the specimen assembly of Fig. 1 was suspended. The furnace, which is shown in Fig. 4, consisted essentially of a graphite cylinder insulated from an external water-cooled steel jacket by a packing of carbon black. Six windows permitted observation of the specimen and electrodes, and optical pyrometer measurement of the specimen temperature. Since the strength of graphite was found to increase with temperature at low temperatures and then to decrease sharply at very high temperatures, two methods of breaking the specimens were utilized. For temperatures above 3000°K, it was found convenient to apply a

FIG. 5. Graphite creep specimen.

load and then raise the temperature until the specimen broke. Below 3000°K, the specimens were heated to various temperatures and then loaded to failure. Unfortunately, no strain measurements were made in these early tests.

B. Creep Characteristics

The creep measurements were made with essentially the same equipment used in the short-time tensile testing. The image of the white-hot specimen was focused on a ground glass screen about five feet from the specimen by means of an achromatic lens in one of the furnace windows. Two reference shoulders on the specimen (as shown in Fig. 5) defined the gauge length used for elongation measurements. Most of the elongation measurements were made upon the screen with a micrometer microscope while the test was in progress. When the time between readings was less than two minutes, however, photographs of the screen were made with a camera using 4×5 in. glass plates. A fixed reference gauge of metal foil was included on the ground glass screen to permit correction for dimensional changes of the photographic emulsion during processing.

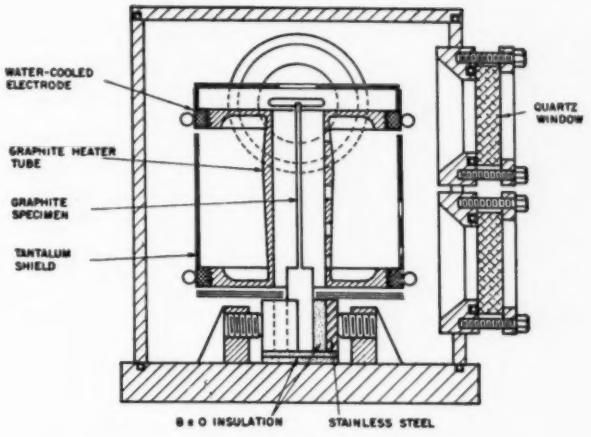


FIG. 6. Cross section of furnace used in Young's modulus determinations.

C. Young's Modulus

The Young's modulus of grade ECA graphite was determined as a function of temperature by measuring the frequency of vibration of specimens fabricated in the form of a cantilever beam. The specimens were heated in a vacuum by radiation from a surrounding cylinder carrying an electric current as is shown in Fig. 6. The experiments consisted of displacing the free end of the specimen, releasing it, and photographing the resulting motion of the beam extremity with a General Radio Company Type 651-AE oscillograph recorder. A typical trace of the incandescent tip obtained in this manner is shown in Fig. 7. The 60-cycle marks along the edge of the film (provided by a spark coil actuated by a General Radio Company Type 631-B Strobotac) furnished a time base which permitted calculation of the frequency and the Young's modulus. No calculation of the logarithmic decrement was made, since it was believed that the support losses would obscure any significant changes in the internal friction of the material. Attempts to measure the modulus dynamically at room temperature by viewing the vibrating specimen with the Strobotac were unsuccessful due to the rapid damping of the vibration, and so a static measurement was substituted.

D. Experimental Errors

The method of heating the specimens in the breaking strength and creep tests leads to uncertainities in the interpretation of the data. Although the core temperature of the specimen was higher than the surface temperature, no correction for this temperature differ-

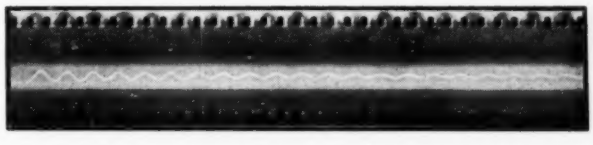


FIG. 7. Initial portion of a typical record obtained in Young's modulus measurement.

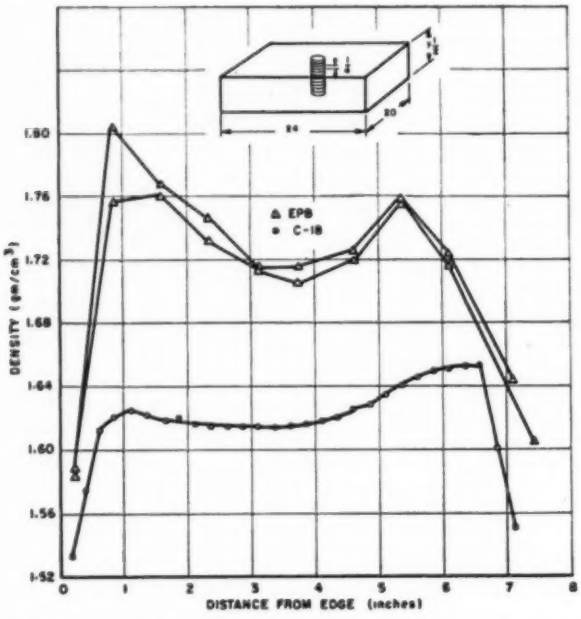


FIG. 8. Density of grades EBP and C-18 graphite as a function of position in molded block.

ence was attempted, and the results are presented in terms of surface temperature. Due to the temperature difference, however, sublimation at the core possibly gave rise to unknown changes in the cross-sectional area of the specimen in the tests performed above 2500°C. In addition, the rapid creep rate found at high temperatures must have had a profound effect upon the strength measurements at those temperatures. In the creep

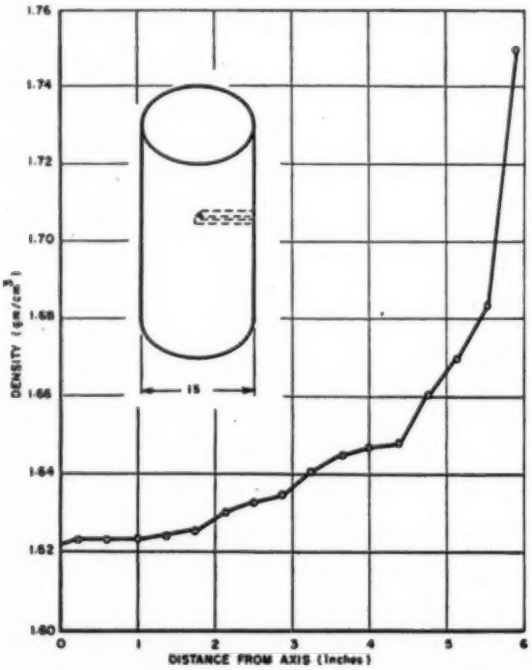


FIG. 9. Density of grade ECA graphite as a function of position in extruded rod.

study, the indirect elongation measurements were subject to errors arising from distortion in the optical system.

The experimental errors involved in the temperature measurements are considered negligible in comparison with the other uncertainties involved. The measurements were made with a disappearing filament optical pyrometer and were corrected for the emissivity of the graphite and the transmissivity of the window.

A possibly significant experimental error in the Young's modulus determinations may be attributed to uncertainties in measurement of the thickness of the cantilever specimen which was only 0.050 in. Since the moment of inertia of the beam section involves the cube of this value, an uncertainty of 0.001 in. can lead to a

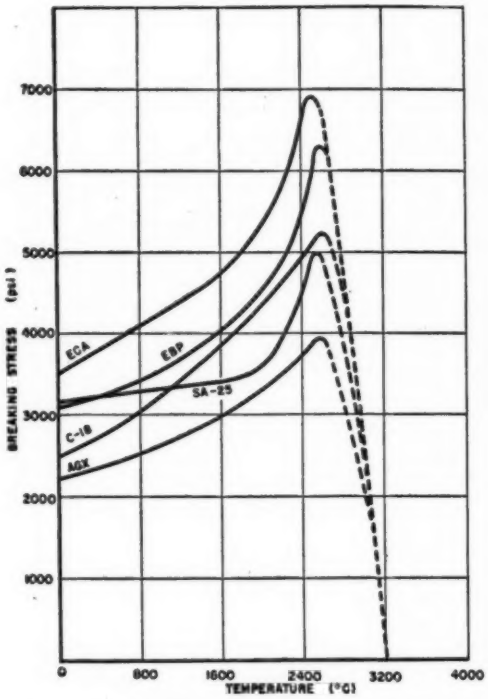


FIG. 10. Short-time breaking strength of various grades of graphite as a function of temperature.

noticeable error in the modulus calculation. Errors of lesser magnitude were contributed by nonuniformity of the temperature distribution along the beam, and by any uncertainties in beam-length measurement and in frequency comparison.

By far the largest error entering into all the graphite experiments, however, arises from the lack of homogeneity of the material. Not only may the presence of voids and fissures in the structure influence the results obtained with the thin specimens used in the modulus study, but variations in the apparent bulk density of the material lead to scatter in the experimental results which probably overshadows most of the errors mentioned above. The bulk density of the molded graphites C-18 and EBP (as measured from disks sawed from a

core taken from the center of the block) is shown in Fig. 8 as a function of positions in the block. A density profile of the extruded ECA graphite rod is presented in Fig. 9.

III. RESULTS AND DISCUSSION

A. Tensile Breaking Strength

The breaking strength of each of the different grades of graphite tested during the present investigation is presented in Fig. 10 as a function of temperature of the specimen. It may be seen that the strength of graphite approximately doubles as the temperature rises from room temperature to about 2500°C. The abrupt decrease in strength observed at higher temperatures may presumably be attributed to the beginning of significant sublimation and of rapid creep.

The smooth curves of Fig. 10 were fitted to experimental points obtained by testing several specimens at each temperature. For all grades but EBP, four specimens were used. Measurements of the apparent (bulk) density of the specimens showed a relatively small variation, and no correlation between density and strength was apparent. The EBP specimens, however, exhibited unusually large variations in both density and strength. In this case, six specimens were tested at each temperature and the data obtained were corrected for density variations.

The large scatter observed in the tensile testing of the EBP specimens is shown in Fig. 11. When the strength values at each temperature were plotted against density (determined by using samples from the cores of the broken tensile specimens), a linear relationship was found. When this relationship was used to correct the observed strength values to an average observed density of 1.67 g/cc, the scatter was reduced to a point where the average values could be approximated by a smooth curve, as is shown in Fig. 12.

No structural explanation for the increase in the breaking strength of graphite with increasing temperature is available at the present time. The temperature to which graphite is subjected during its manufacture is stated[1] to be between 2500° and 3000°C, so that it appears doubtful that any further graphitization can occur. On the assumption that the relief of internal stresses or changes in grain orientation might take place at high temperature, many specimens were heated (with and without applied stress) and then cooled to room temperature. No significant difference between the room-temperature strength of these samples and that of untreated graphite was found.

X-ray diffraction studies[3] were made on several specimens which had been broken at high temperature. Grain orientations parallel and perpendicular to the load direction were examined. Within the limit of accuracy of the method (0.1 percent) no change was found in the orientation or the average particle size in these specimens, as compared with unstrained graphite.

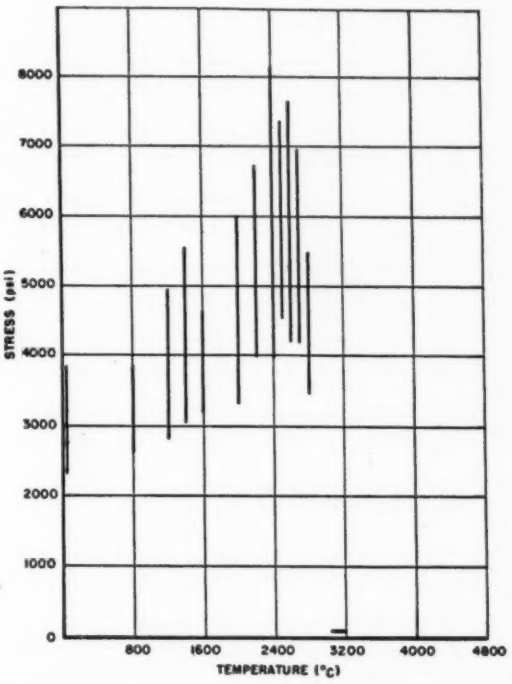


FIG. 11. Range of variation of values obtained in strength measurements with EBP graphite.

A maximum in the tensile breaking strength *versus* temperature curve has been reported for a sillimanite refractory at 980°C.[4] Results obtained with tin and

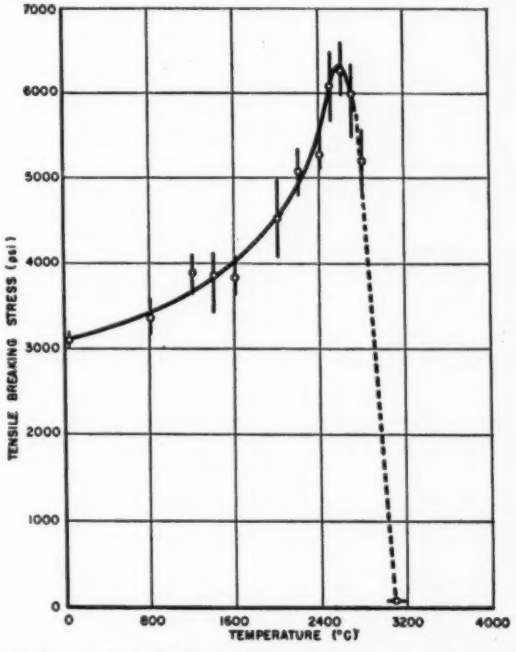


FIG. 12. Strength of EBP graphite after correcting to a density of 1.67 g/cc.

[3] Work performed by Dr. Ralph Hultgren, University of California, Berkeley. Earlier work with same results was done by Dr. Pol Duwez, California Institute of Technology.

[4] Kunen, Hartwig, and Bressman, "Tensile properties of a sillimanite refractory at elevated temperatures," NACA Tech. Note No. 1165 (1946).

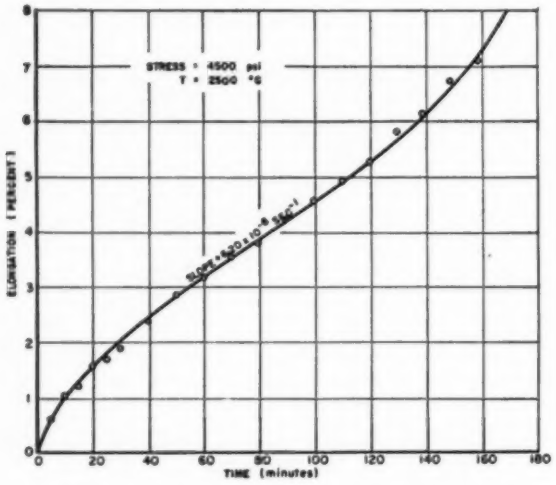


FIG. 13. Typical creep curve for grade ECA graphite.

some tin alloys indicate similar behavior at temperatures in the range of −160 to −60°C,[5,6] although the data are presented in terms of ultimate strength and area reduction. The characteristics exhibited by graphite, however, differ from those of the above materials in that the increase of short-time strength with temperature, as shown by Fig. 8, appears to be limited only by the imminence of sublimation. If corrections could be made for the large influence of creep and for the decrease in effective specimen cross section due to sublimation of material at the hot core, the abrupt drop in short-time strength might possibly appear even more precipitous.

It may be of interest to note that of the various fracture criteria† proposed in the literature, the energy relationship advanced by Saibel[7] shows promise of giving the most satisfactory qualitative agreement with the behavior exhibited by graphite. Saibel's criterion, which relates fracture to the phenomenon of melting, is based upon the assumption that "the quantity of energy required for the abolition of cohesive strength is equal to the fractional change in specific volume in passing from the solid to the liquid state multiplied by the heat of fusion."

Until Saibel's argument concerning the formation of "holes" can be clarified and extended to include phase changes from the solid to the gaseous state, however, and until stress-strain data for graphite are available, quantitative speculation concerning an energy criterion for the failure of graphite will not be justified. It is suggested that the interpretation of graphite behavior might possibly be aided by strength *versus* temperature measurements upon other materials (such as possibly

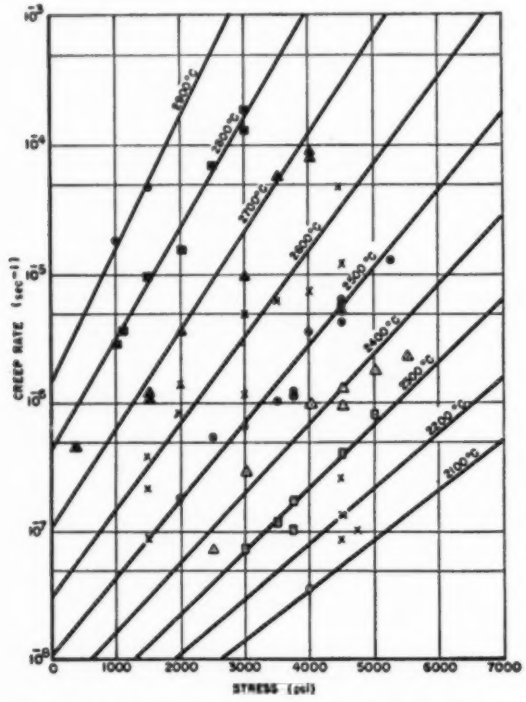


FIG. 14. Steady creep rate obtained with ECA graphite as a function of stress for various temperatures.

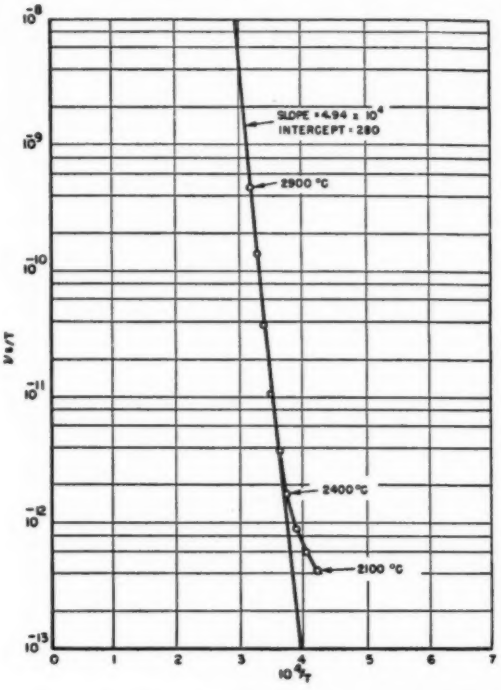


FIG. 15. Determination of activation energy and entropy change for ECA graphite.

[5] H. S. Kalish and F. J. Dunkerley, "The low temperature properties of tin and tin-lead alloys," AIME Tech. Pub. No. 2442, (1948).

[6] Dunkerley, Hunter, and Stone, Metal Trans. 185, 1005 (1949).

† Comprehensive surveys have been made by Orowan (E. Orowan, "Fracture and strength of solids," contained in *Reports on Progress in Physics* (The Physical Society, 1948–49), Vol. XII); and Gensamer, *et al.* (Gensamer, Saibel, Ransom, and Lowrie, *The Fracture of Metals* (American Welding Society, 1947).)

[7] E. A. Saibel, "A thermodynamic theory of the fracture of metals," Metals Technology, T.P. 2131 (February, 1947).

iodine) in a thermodynamic state (relative to their triple points) where melting would not occur.

B. Creep Characteristics

The curves of elongation *versus* time which were obtained with graphite are qualitatively similar to those characteristic of metals. A typical curve is presented in Fig. 13. The slope of the apparently linear portion of this curve defines the rate of so-called secondary steady-state creep. The experimental data obtained with ECA graphite at various stresses and temperatures are presented in Fig. 14 in terms of this steady creep rate, a form of presentation which has become conventional.

The experimental data as presented in Fig. 14 may be interpreted in terms of the rate theory evolved by Eyring[10] and utilized by Kauzmann,[11] Dushman, *et al.*,[12] and others. Such theory as applied to a shear process defines an activation energy and an entropy change which are related by the following equation:

$$\log_{10}\left(\frac{\nu_s}{T}\right)=\log_{10}\left(\frac{k}{h}\right)+\frac{\Delta\eta}{2.303R}-\frac{Q}{2.303RT}, \quad (1)$$

where ν_s = zero stress intercept of constant temperature lines on graph of creep rate *versus* stress, T = absolute temperature, R = gas constant, k = Boltzmann's constant, h = Planck's constant, Q = activation energy, $\Delta\eta$ = change in entropy. When the graphite data are plotted in terms of $\log_{10}(\nu_s/T)$ *versus* $(1/T)$ as shown in Fig. 15, the slope and intercept of the "straight" portion of the curve determine the activation energy and entropy change to be

$$Q=226{,}000 \text{ cal/mole},$$
$$\Delta\eta=-174 \text{ cal/mole } °\text{K}. \quad (2)$$

The deviation of the data presented in Fig. 15 from a straight line may be interpreted as an indication that the rate of creep may be governed by two different mechanisms, one predominant at high temperatures, the other prevailing in the low temperature regime. The lack of sufficient data, however, precluded calculation of the energy and entropy values for the lower temperature range. Speculation regarding the existence of a change in crystal structure at about 2500°C is encouraged by a rapid decrease in the time required for establishment of a "linear" creep rate which was noted at this temperature, and by the maximum observed in the tensile strength curve at the same temperature.

Deductions concerning the behavior of graphite based upon the Eyring-Kauzmann theory, however, cannot be accepted without reservation. In the first place, as has been suggested by Lubahn[13] and others, the distinction between secondary, steady-state creep (upon which the rate theory is predicated) and primary, transient creep may be illusory. One of the factors contributing to the steady-state illusion is the so-called tertiary creep (as exhibited by the curve of Fig. 13), which in the present work is attributed to reduction of the specimen cross section, reduction aggravated at high temperatures due to loss of material by sublimation. Secondly, the rate theory was derived for a process of pure shear, and graphite exhibits some properties which appear difficult to explain by a shear mechanism. The fracture surfaces obtained with graphite specimen at all temperatures up to the sublimation point, for example, are typical of the pure tensile fracture of a brittle material even though the failure might have been preceded by a considerable

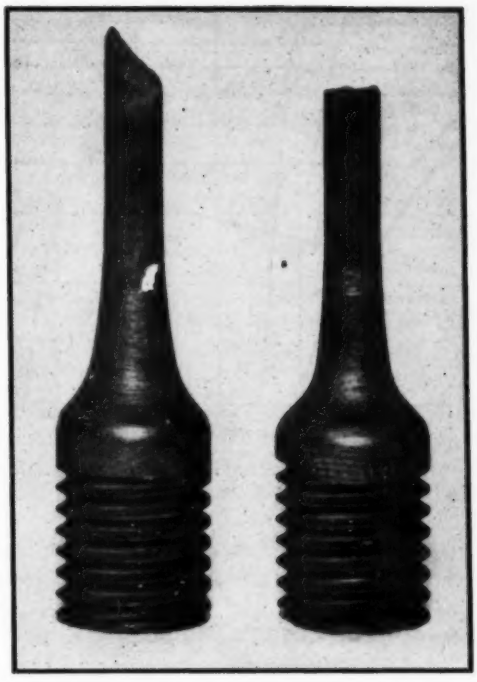


FIG. 16. Typical fractures obtained in torsion and tension tests at 2700°C. Left: torsion. Right: tension.

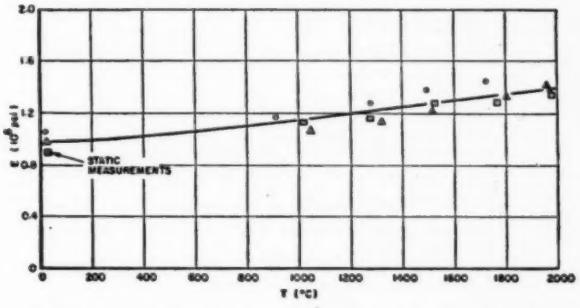


FIG. 17. Young's modulus of ECA graphite as a function of temperature.

[10] H. Eyring, J. Chem. Phys. 4, 283 (1936).
[11] W. Kauzmann, "Flow of solid metals from the standpoint of chemical-rate theory," AIME Tech. Pub. No. 1301.
[12] Dushman, Dunbar, and Huthsteiner, J. Appl. Phys. 15, 108 (1944).
[13] J. D. Lubahn, "Creep of metals," contained in a seminar, *Cold working of metals* (American Society for Metals, 1949).

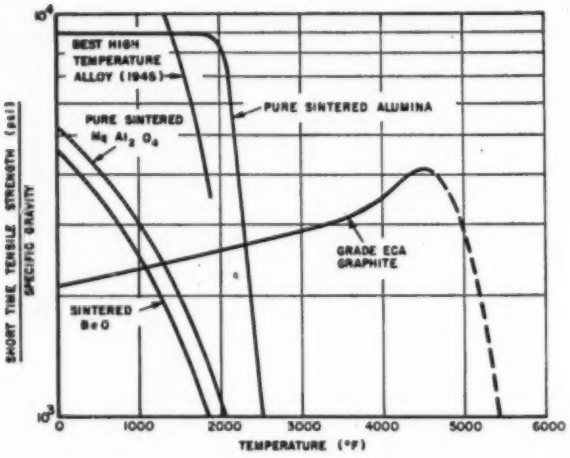


FIG. 18. Strength-to-weight ratio for several materials as a function of temperature.

amount of permanent elongation.‡ Typical fractures obtained in torsion and tension are shown in Fig. 16.

C. Young's Modulus

The Young's modulus of three specimens of grade ECA graphite is presented as a function of temperature in the range from 1000 to 2000°C in Fig. 17. Temperature variations along the specimen length (of the order of 10 percent at the greatest) required presentation of the results in terms of an average temperature, an average determined, in this case, by using the nominal fiber strain in the specimen as a weight function. Figure 17 also shows values of the modulus which were obtained by measurement of static deflections of the same specimens at room temperature. Considering the brittleness of graphite in the above temperature range, it is believed that the modulus measured dynamically at the low frequency employed (about 70 cycles/sec) should not differ materially from the static modulus.

The increase in the Young's modulus with increasing temperature observed for ECA graphite cannot be explained at the present time. In spite of the scatter due to the possible experimental errors discussed earlier, the results shown in Fig. 17 are regarded as qualitatively correct. It seems apparent that the thermal dependence of the macroscopic elasticity coefficients of a pure solid[14] may be influenced by the state of the material relative to its triple point. It should be mentioned that preliminary torsion experiments have indicated the existence of a possible maximum point (at about 2100°C) in the modulus of rigidity of ECA graphite which represents an increase of about 20 percent over the room temperature value. In this case, of course, interpretation is complicated by the fact that the shearing stress was in the plane of minimum tensile strength, normal to the direction of extrusion. The directional dependence of the shear properties of graphite has not yet been investigated, but it is presumed that the maximum shear resistance would be found in this plane. The torsion work is continuing and will be reported in detail at a later date.

‡ Consideration of the porous nature of graphite may possibly permit reconciliation of the apparently brittle nature of the gross fracture surface with plastic deformation of the individual elements of the porous structure.

[14] L. Brillouin, Phys. Rev. **54**, 916 (1938).

IV. CONCLUSION

The experiments described above represent a preliminary investigation of some pertinent mechanical properties of graphite which was made in order to evaluate the suitability of the material for high temperature structural applications. It was found that the tensile-breaking strength of the several grades of graphite considered assumes its maximum value at a temperature of about 2500°C. The strength measurements in this region were profoundly influenced, however, by the high creep rate which was observed and investigated at corresponding temperatures. The Young's modulus of grade ECA graphite was found to increase slowly with temperature, its value at 2000°C being about 40 percent greater than its room temperature value.

The unusual high temperature strength properties and low density exhibited by graphite make it appear promising as a potential structural material. In order to facilitate comparison of the characteristics of graphite with those of some other materials proposed for gas turbine application, for example, the tensile data for ECA graphite are presented in Fig. 18 together with corresponding data for several ceramic oxides and a good high temperature alloy (as reported by Norton[15]) in terms of the short-time tensile strength divided by the nominal specific gravity of the material. It is apparent that for the same creep rates, graphite can be used at temperatures much higher than those permissible with other materials.

The unique behavior which was observed in the present investigation emphasizes the desirability of further study of the properties of graphite at elevated temperatures. A need for fatigue data and accurate stress-strain information obtained under controlled conditions of strain rate and temperature is apparent, and measurements of electrical resistivity and internal friction might provide useful information in regard to possible changes in grain orientation and crystal structure.

ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of W. Danielson, A. Gorton, and H. Strahl, whose work made possible the early completion of the experimental program.

[15] F. H. Norton, *Refractories* (McGraw-Hill Book Company, Inc., New York, 1949), third edition, p. 664.

Electromagnetic Scattering from Spheres with Sizes Comparable to the Wavelength*

A. L. ADEN†
Geophysical Research Directorate, Air Force Cambridge Research Laboratories, Cambridge, Massachusetts
(Received October 26, 1950)

By using suitably defined logarithmic derivative functions and the recurrence formulas due to Infeld, the problem of obtaining numerical answers for the scattering of a plane electromagnetic wave by a sphere can be simplified. This is particularly true if the sphere has a complex index of refraction. The nature of these logarithmic derivative functions is shown for one special case involving water spheres.

Experimental results are given for the back-scattering cross sections of water spheres in the electrical size region $0.74 \leq 2\pi a/\lambda \leq 5.90$. These results were obtained at a wavelength of 16.230 cm using the standing-wave method of measurement. It is believed that these were the first laboratory measurements of the back-scattering cross section made on individual water spheres in this critical size region.

Comparison is made between the experimental and theoretical results.

I. INTRODUCTION

THIS paper deals with the determination of the back-scattering cross section for spheres with sizes comparable to the wavelength. Particular attention is directed toward the case of water spheres for which the index of refraction is complex.

A. The Meaning of the Back-Scattering Cross Section

The back-scattering cross section σ is a lumped measure of the ability of a scattering obstacle to re-radiate energy in the direction of the source. Under the assumption of an incident plane electromagnetic wave at the scatterer, this parameter can be determined from a knowledge of the far zone scattered field. The defining relation is

$$\sigma = 4\pi r^2 (E_s/E_0)^2, \tag{1}$$

where E_0 is the magnitude of the incident electric field at the scattering object, and E_s is the magnitude of the far zone scattered field at a distance r from the scatterer in the direction of the source.

B. The Problem of Microwave Reflection from Rain

With the development of microwave radar during the war, it was found that for sufficiently small wavelengths rain could produce an appreciable echo. This echo phenomenon was important for several reasons. For the usual tactical use of the equipment, it was important to distinguish between atmospheric reflections and operational targets. Also, as a meteorological tool, the observance of echoes was very useful in mapping the rain areas and in showing their movements.

In order to investigate theoretically microwave reflection from rain, it is necessary to assume that the drops are randomly distributed in space, and that the mutual interaction between drops is negligible. The first assumption appears to be valid without further comment. The second assumption is based on Trinks'[1] analysis that for Rayleigh scattering the mutual interaction is negligible for sphere spacings greater than two or three sphere diameters, plus the fact that in actual rain the average spacing between drops is many times this value.[2] Under these conditions the problem can be divided into two parts: (1) finding the drop size distribution, and (2) finding the reflection from a single drop. The drop size distribution is strictly a meteorological problem; the reflection from a single drop is a problem in electrodynamics.

The first thorough investigation of the reflection and attenuation effects from rain was made by Ryde *et al.*[3–7] Ryde's method was to compute the numerical results for a single water sphere using the classical Mie[8] theory and to use the assumptions given above to make application to the case of many drops. Further contributions to this problem have been made by Goldstein,[9] Haddock,[10] and others. In all cases, a major difficulty has been the evaluation of the reflection from a single sphere. It was an interest in this problem that prompted the present investigation.

II. THEORETICAL DETERMINATION OF THE BACK-SCATTERING CROSS SECTION

A. The Classical Solution

The first thorough treatment of the problem of the scattering of a plane electromagnetic wave by a sphere

* The research reported in this paper was performed while the author was an RCA Fellow in electronics under the National Research Council. Additional support was extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force under ONR contract N5 ori-76, T.O.1.

† Formerly, Cruft Laboratory, Harvard University, Cambridge, Massachusetts.

[1] W. Trinks, Ann. Physik **22**, 561 (1935).
[2] W. Palmer, "Studies of continuous precipitation," Doctoral thesis, McGill University, Montreal, Canada, April, 1949.
[3] J. Ryde, G. E. Co. Report No. 7831 (1941).
[4] J. Ryde and D. Ryde, G. E. Company Report No. 8516 (1944).
[5] J. Ryde and D. Ryde, G. E. Company Report No. 8670 (1945).
[6] J. Ryde, *Meteorological Factors in Radio-Wave Propagation* (The Physical Society, 1 Lowther Gardens, Prince Consort Road, London, 1946), pp. 169–189.
[7] J. Ryde, J. Inst. Elec. Engrs. (London) **93**, 101 (1946).
[8] G. Mie, Ann. Physik **25**, 377 (1908).
[9] L. Goldstein, Off. Pud. Bd., Report PB 5850, NDRC Report WPG-11 (1945); also published in *Radio Wave Propagation* (Academic Press, Inc., New York, 1949), Vol. II, Chapter 5.
[10] F. Haddock, paper presented at the U.R.S.I.—I.R.E. meeting, Washington, D. C. (1947).

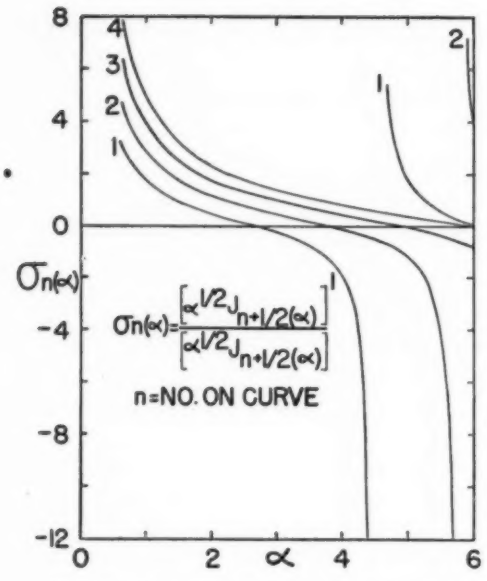


FIG. 1. The σ_n function for real argument.

of arbitrary size and any electric properties was given by Mie[8] in 1908. The present approach is to use the method of Stratton[11] who has shown that the complete theory can be given neatly and concisely by the use of orthogonal spherical vector wave functions.

Using the results of Stratton,‡ the back-scattering cross section, normalized with respect to the geometrical cross section of the sphere, is given by

$$\frac{\sigma}{\pi a^2}=(1/\alpha^2)\left|\sum_{n=1}^{\infty}(2n+1)(-1)^n(a_n^s-b_n^s)\right|^2. \quad (2)$$

Here, $\alpha=2\pi a/\lambda$, where a is the radius of the sphere and λ is the wavelength. a_n^s and b_n^s are the scattering amplitude coefficients given by

$$a_n^s=-\frac{j_n(N\alpha)[\alpha j_n(\alpha)]'-j_n(\alpha)[N\alpha j_n(N\alpha)]'}{j_n(N\alpha)[\alpha h_n^{(2)}(\alpha)]'-h_n^{(2)}(\alpha)[N\alpha j_n(N\alpha)]'}, \quad (3)$$

$$b_n^s=-\frac{j_n(\alpha)[N\alpha j_n(N\alpha)]'-N^2 j_n(N\alpha)[\alpha j_n(\alpha)]'}{h_n^{(2)}(\alpha)[N\alpha j_n(N\alpha)]'-N^2 j_n(N\alpha)[\alpha h_n^{(2)}(\alpha)]'}. \quad (4)$$

The primes at the square brackets indicate differentiation with respect to the argument of the bessel function inside the brackets. N is the index of refraction which may be real or complex. $j_n(\alpha)$ is the spherical bessel function defined by

$$j_n(\alpha)=(\pi/2\alpha)^{1/2}J_{n+1/2}(\alpha) \quad (5)$$

and $h_n^{(2)}(\alpha)$ is the spherical hankel function of the second kind defined by

$$h_n^{(2)}(\alpha)=(\pi/2\alpha)^{1/2}H_{n+1/2}^{(2)}(\alpha). \quad (6)$$

B. The Problem of Evaluation

The equations given above represent an elegant formal solution to the problem. It is seen that the evaluation of the back-scattering cross section reduces to the determination of the scattering amplitude coefficients and the summation of the appropriate series. However, the task of actually carrying out this process is usually not easy. Examination of Eqs. (3) and (4) reveals the difficulty: both a_n^s and b_n^s are complicated expressions involving bessel and hankel functions and their derivatives, and, even assuming N is real, the problem of evaluating them is long and tedious. If N is complex, the task of evaluating Eqs. (3) and (4) in their present form becomes almost impossible. The reason for this is that no extensive tables are available for the spherical bessel functions of complex arguments.

The difficulties mentioned above can be overcome in certain special cases by using approximations. For example,[11] the case of highly conducting spheres, and the limiting cases of very large or very small spheres can be so handled. However, in many practical problems one is interested in the general case of spheres with sizes comparable to the wavelength and with a complex index of refraction. In this case, one is faced with the task of evaluating the general equations for orders up to $n\sim 2\alpha$. Fortunately, this is now possible.

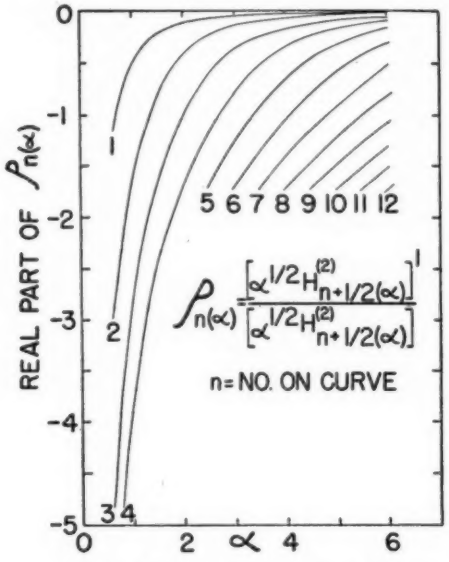


FIG. 2. Real part of the ρ_n function.

C. The Method of Logarithmic Derivative Functions

The method to be described consists of transforming the equations for a_n^s and b_n^s by using suitably defined logarithmic derivative functions and showing that these logarithmic derivative functions can be evaluated. This

[11] J. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Chapter IX, Sections 9.25, 9.26, and 9.27.

‡ $e^{+j\omega t}$ time dependence is assumed here, instead of $e^{-i\omega t}$ assumed by Stratton.

method was first used by Infeld[12] in connection with antenna problems. It has also been used in this connection by Smith[13] and Tai.[14] The notation used here is due to Tai.

If one defines

$$\rho_n(x)=\frac{d}{dx}\ln[xh_n^{(2)}(x)]=\frac{d}{dx}\ln[x^{(1/2)}H_{n+1/2}^{(2)}(x)], \quad (7)$$

$$\sigma_n(x)=\frac{d}{dx}\ln[xj_n(x)]=\frac{d}{dx}\ln[x^{(1/2)}J_{n+1/2}(x)], \quad (8)$$

then

$$a_n^s=-\frac{j_n(\alpha)}{h_n^{(2)}(\alpha)}\left\{\frac{\sigma_n(\alpha)-N\sigma_n(N\alpha)}{\rho_n(\alpha)-N\sigma_n(N\alpha)}\right\}, \quad (9)$$

$$b_n^s=-\frac{j_n(\alpha)}{h_n^{(2)}(\alpha)}\left\{\frac{\sigma_n(N\alpha)-N\sigma_n(\alpha)}{\sigma_n(N\alpha)-N\rho_n(\alpha)}\right\}. \quad (10)$$

Equations (9) and (10) appear to be simpler than Eqs. (3) and (4), since there are fewer functions present and no derivatives involved explicitly. However, it still remains to be shown how the functions involved here can be evaluated. This will now be done.

Since α is always real, the ratio $j_n(\alpha)/h_n^{(2)}(\alpha)$ can be evaluated using standard tables for the spherical functions. The logarithmic derivative functions of real arguments can also be computed directly using the defining equations, (7) and (8). However, since these equations require the evaluation of the derivatives of the spherical functions, it is simpler to use equivalent forms in which the derivatives do not appear explicitly. These are

$$\rho_n(\alpha)=\frac{h_{n-1}^{(2)}(\alpha)}{h_n^{(2)}(\alpha)}-\frac{n}{\alpha}, \quad (11)$$

$$\sigma_n(\alpha)=\frac{j_{n-1}(\alpha)}{j_n(\alpha)}-\frac{n}{\alpha}. \quad (12)$$

Using Eqs. (11) and (12), the logarithmic derivative functions of real arguments can be evaluated using standard tables for the spherical functions. The real difficulty comes again when the functions involve complex arguments. Equations (11) and (12) cannot be used then because there are no extensive tables available for the spherical functions of complex arguments. Here one makes use of the recurrence formula derived by Infeld.[12] This formula is

$$\sigma_n(x)=\frac{x^2+nx\sigma_{n-1}(x)-n^2}{nx-x^2\sigma_{n-1}(x)}. \quad (13)$$

By use of this recurrence formula, the σ_n function of any order can be computed provided the next lower order function is known. However, the zeroth-order term can be computed directly; i.e.,

$$[xj_0(x)]=\sin x, \quad (14)$$

$$[xj_0(x)]'=\cos x, \quad (15)$$

so that

$$\sigma_0(x)=\cot x. \quad (16)$$

Hence, if x is complex, say $x=c-jd$,

$$\sigma_0(x)=\cot(c-jd)=\frac{\sin 2c+j\sinh 2d}{\cosh 2d-\cos 2c}. \quad (17)$$

Higher order terms can now be found using Eq. (13).

Thus, by writing the classical solution in terms of suitably defined logarithmic derivative functions, the scattering amplitude coefficients can be computed exactly, even in regions where the spherical bessel functions of complex argument are not tabulated.

D. The Nature of the Logarithmic Derivative Functions

Using the procedure outlined above, computations were made for the logarithmic derivative functions $\rho_n(\alpha)$, $\sigma_n(\alpha)$, and $\sigma_n(N\alpha)$ over the range $0.6<\alpha<6$. The value of N used was that for the index of refraction of water at the wavelength of 16.230 cm which was used in the experimental measurements described in the next section.

Figure 1 shows the σ_n function for real argument plotted against the electrical sphere size α for $n=1,2,3,4$. It is seen that $\sigma_n(\alpha)$ is a smooth non-oscillating function. It approaches $+\infty$ as $(n+1)/\alpha$ when α approaches zero. It has poles at the zeros of $J_{n+1/2}(\alpha)$. However, the poles of $\sigma_n(\alpha)$ do not cause the scattering amplitude

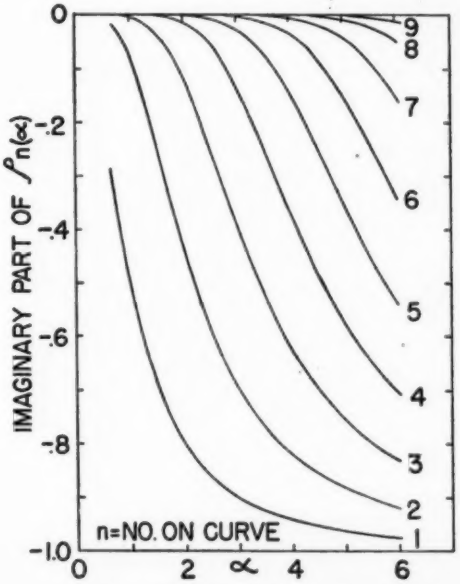


FIG. 3. Imaginary part of the ρ_n function.

[12] L. Infeld, Quart. Appl. Math. **5**, 113 (1947).
[13] P. Smith, J. Appl. Phys. **19**, 11 (1948).
[14] C. Tai, Report No. 77, Cruft Laboratory, Harvard University (1949).

coefficients to blow up, since it is always the product $\sigma_n(\alpha) \cdot J_{n+1/2}(\alpha)$ which appears in Eqs. (9) and (10), and this product has no poles. For higher orders, the curves are similar in shape to those shown.

Figures 2 and 3 show the real and imaginary parts of $\rho_n(\alpha)$. These functions are also non-oscillatory. The real part approaches $-\infty$ as $-n/\alpha$ when α approaches zero. The imaginary part starts at zero and approaches -1 as α increases.

Figure 4 shows the real and imaginary parts of $\sigma_1(N\alpha)$. It is seen that these are rapidly oscillating, highly damped functions. The real part approaches zero and the imaginary part approaches 1 as α increases. Higher order curves are similar in behavior.

III. EXPERIMENTAL DETERMINATION OF THE BACK-SCATTERING FROM WATER SPHERES

A. Statement of the Problem

The experimental measurement of the reflection from individual water spheres involves several important difficulties: (1) The sphere is a very low gain reflector, and this makes it hard to get reliable answers by ordinary pulsing techniques. (2) It is not easy to maintain a water sphere while measurements are being made. For these reasons, no direct laboratory measurements on individual water spheres have previously been made successfully. Even now, these difficulties almost preclude the taking of measurements in the region of greatest interest, that is in the spectral region where water spheres of actual raindrop size are comparable to the wavelength. However, it is possible to perform an experiment at slightly longer wavelengths and to increase the size of the water spheres to get into the critical size region. This is the experiment actually performed. It is believed that this technique is useful, since comparison between experiment and theory at one frequency should be a good indicator of expected comparisons at other frequencies.

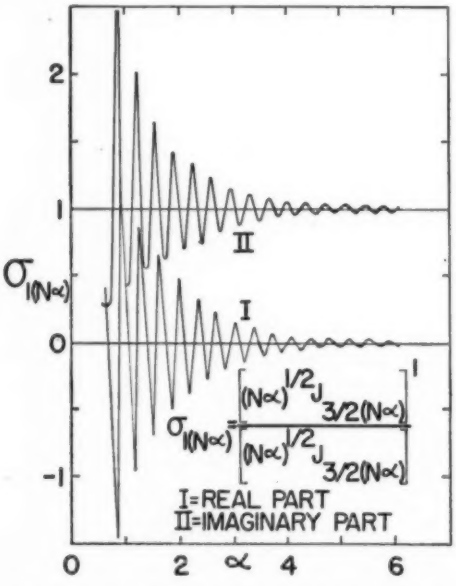


FIG. 4. The σ_1 function for complex argument.

B. The Experimental Techniques

The experimental method of measurement used in the present investigation was the standing-wave method of D. King.[15] This method utilizes an image screen technique, and the standing waves set up on the screen by the interaction of the incident and reradiated waves are measured along the radial line between the scatterer and source. This method offers the advantages of a system having absolute calibration, relatively simple equipment, and measurements at low power. The main disadvantages are that it requires an image screen which is large, uniform and rigid, and that the obtainment of results is very time consuming. The derivation of the formulas needed and a discussion of the approximations, as well as a detailed description of the equipment used, is given elsewhere.[16]

One of the worst obstacles to making measurements on water spheres is the inability to maintain such a sphere while measurements are being taken. This obstacle was overcome by using as a container thin hemispherical shell forms of Styrofoam§ which were mounted on aluminum disks. Since Styrofoam has dielectric properties extremely close to those of air, it had a negligible effect on the measurements. Thus, when the form was filled with water and the disk was inserted into its proper place in the image screen, the effect was a hemisphere of water exposed over a large image screen. By image theory, the measurements taken were the same as for a complete sphere in free space.

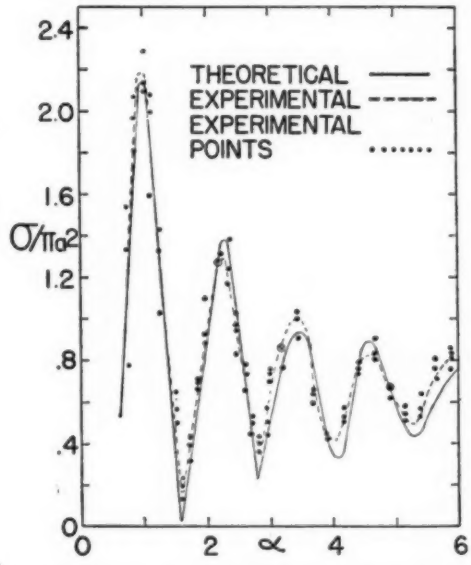


FIG. 5. Back-scattering from water spheres.

[15] D. King, Technical Report No. 50, Cruft Laboratory, Harvard University (1948); Proc. Inst. Radio Engrs. **37**, 770 (1949).

[16] A. Aden, Technical Report No. 106, Cruft Laboratory, Harvard University (1950).

§ Product of The Dow Chemical Company, Midland, Michigan.

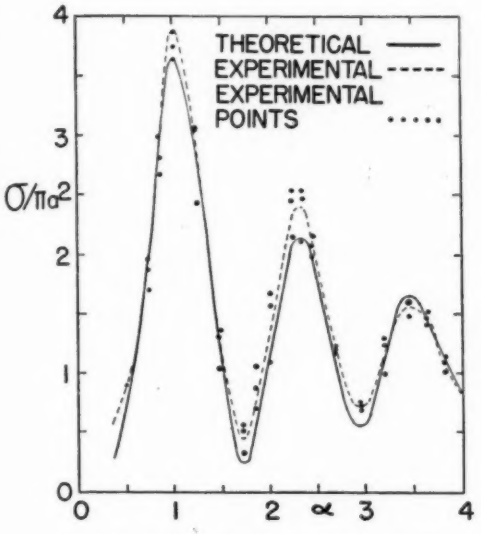


FIG. 6. Back-scattering from metal spheres.

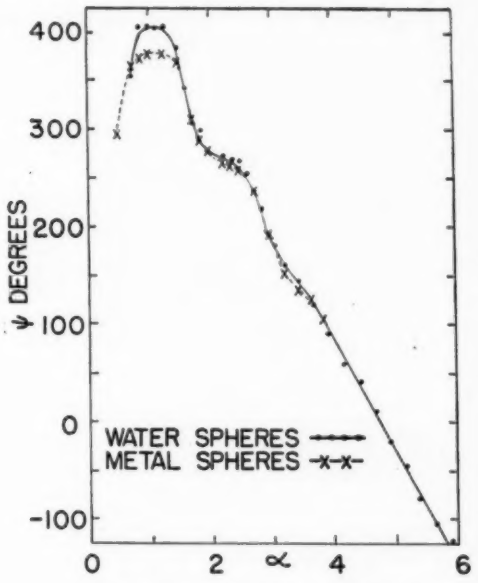


FIG. 7. Phase shift of the E field.

C. The Experimental Data

Measurements of the back-scattering cross section were made on thirty water spheres in the electrical size region $0.74 \leq \alpha \leq 5.90$. It is believed that these measurements were the first laboratory measurements made on individual water spheres in this critical size region.

As a check on the standing-wave method, measurements of the back-scattering cross section were made on eighteen metal spheres. The metal spheres served as good standard scatterers since experimental verification had already been made by Yates[17] using another method of measurement.

Finally, the phase shift ψ of the far zone electric field at the scattering spheres has been determined. This data comes directly from a knowledge of the positions of the extreme values of the standing-wave curves.

IV. COMPARISON OF THE RESULTS

Figure 5 shows a comparison of the theoretical and experimental results for the normalized back-scattering cross section for water spheres at a wavelength of 16.230 cm. The theoretical curve was obtained using the method of logarithmic derivative functions outlined in Sec. II; the experimental results were obtained using the standing-wave method as outlined in Sec. III. It is seen that there is very good agreement between the experimental and theoretical results. For spheres with sizes very much smaller than those considered here, the results would be given to a good approximation by the Rayleigh law. For spheres with sizes larger than those considered here, the results should approach the large size approximation of about 0.64.

Figure 6 shows a comparison of the theoretical[18] and experimental results for the normalized back-scattering cross section for metal spheres. Again, it is seen that there is a good agreement between the experimental and theoretical results. It is interesting to note, in comparing Figs. 5 and 6, that the extreme values of $\sigma/\pi a^2$ occur at approximately the same values of α for both the metal and water spheres. The maximum values on the water sphere curve are roughly 60 percent of the corresponding values on the metal sphere curve.

The phase shift ψ for both the metal and water spheres is plotted in Fig. 7. It is seen that the phase shift is nearly the same for the metal and water spheres for $\alpha > 1.6$. There is a peak in the phase shift curves near the first peak in the normalized back-scattering curves. There are inflection points in the phase shift curves near the second and third peaks in the normalized back-scattering curves. The phase shift is nearly linear for $\alpha > 3.7$.

ACKNOWLEDGMENTS

The author gratefully acknowledges the guidance and encouragement of R. W. P. King during the course of the investigation. Thanks are due Miss J. Klimas and Mrs. E. Haurwitz for assisting with the numerical computations, and to N. C. Gerson for reading the manuscript.

[17] K. Yates, The Ohio State University Research Foundation Report 759–33 (1945).

[18] G. Sinclair (privately communicated).

Electrode Reactions in the Glow Discharge

F. E. HAWORTH
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received November 1, 1950)

The reactions which occur at silver electrodes in a *normal* glow discharge in air have been determined. These are: (1) formation of $AgNO_2$ and some Ag_2O at the anode at the rate of 3.4 μg/coulomb; (2) loss of metal from the cathode by chemical action at the rate of 3.5 μg/coulomb (probably the same reaction as (1) with subsequent loss of the reaction products by the greater heating of the cathode, but this hypothesis has not been established); and (3) normal sputtering loss at the cathode at the rate of 0.4 μg/coulomb. These processes result in building a conducting layer on the anode. If the electrode separation is so small that the anode extends into the region of the cathode fall, then the high electric field pulls the newly formed and not very coherent growth upon the anode across into a bridge between the electrodes.

SINCE Faraday's description of the glow discharge in 1838, it has been the subject of an immense amount of work, most of which has been concerned with the phenomena occurring in the gas between the electrodes. The experiments[1] reported here are concerned with the effect of the glow discharge upon the electrodes themselves, which has had little attention in the past (except for cathode sputtering in the high voltage abnormal glow). Furthermore, the present experiments are limited to the *normal* glow and, since the results are fairly complete only for silver, passing comments only will be made on work carried out upon other metals.

The discharge circuit consists of the electrodes, a variable dc voltage source, and a series resistance. The electrodes are mounted on micrometers for accurate adjustment of the gap, and placed inside a chamber which can be evacuated or filled with suitable gases in order to provide control of the atmosphere. For visual observation of the discharge a microscope having a magnification of 125 is provided.

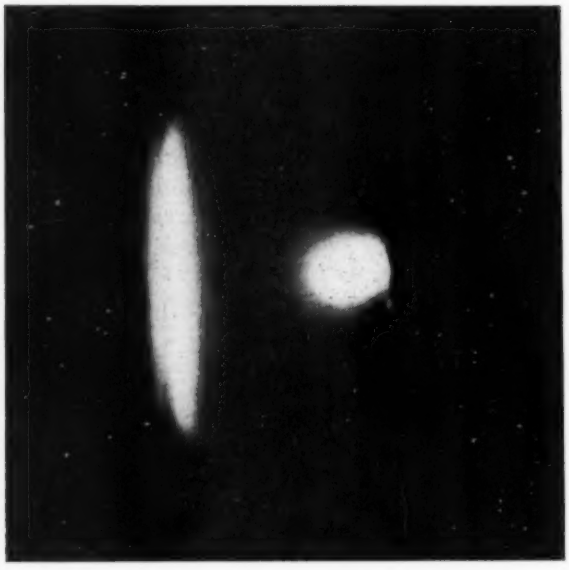


FIG. 1. Photograph of the glow discharge in air at atmospheric pressure. Electrode separation 250×10^{-4} cm and current 0.010 amp. Negative glow at left and positive column at right $\times 108$.

[1] These experiments were undertaken in connection with studies of contact erosion. A. M. Curtis [Bell System Tech. J. **19**, 40 (1940)] has shown that frequently a glow discharge is formed between opening relay contacts.

The electrodes (used for all the experiments except the group of quantitative measurements described later) are small rods 0.127 cm in diameter and about 1.5 cm long. They are mounted end to end, the discharge thus taking place between two opposed ends, and these are slightly rounded in order to keep the discharges centered. The surfaces are prepared by polishing, the final polish being on silk with levigated alumina. The electrode separations are small, that being the condition of greatest interest to us.

GLOW DISCHARGE CHARACTERISTICS AT ATMOSPHERIC PRESSURE

Figure 1 is a photograph of a glow discharge in air at atmospheric pressure with a current of 0.010 amp and a gap of 250×10^{-4} cm. At the left is the bluish white negative glow, about 50×10^{-4} cm thick, reflection from the polished surface making it appear thicker. (The current density in the negative glow is about 9 amp/cm^2 and at 0.110 amp covers the cathode completely.) At the right is the pink positive column, and in between the faraday dark space. The cathode dark space, seen in discharges at low pressure, is too thin to be seen. As the gap is decreased the positive column disappears, and then the faraday dark space until at a 25×10^{-4}-cm gap the anode is immersed in the negative glow.

The potential of the discharge varies little with gap length or current, but changes considerably with electrode material and gas atmosphere. It is about 25 percent lower for nitrogen than for air. With a 25×10^{-4}-cm gap and 0.010-amp current the values in these experiments range from 218 volts for Cu electrodes in nitrogen to 350 volts for W electrodes in air. For Ag electrodes in air the potential is about 300 volts.

ANODE GROWTHS

The most striking effect that the glow produces on the electrodes is found with silver in air when the anode

is immersed in the negative glow at a gap of 25×10^{-4} cm. Under these conditions the discharge continues for only a few seconds and then is suddenly shorted out by a solid bridge across the gap. Such a bridge appears so quickly that its growth cannot be observed. These bridges are about 10×10^{-4} cm in diameter and look like metallic silver. At a gap only twice as great, however, one can observe the process for several minutes before the bridging occurs. A layer of material is seen to form on the anode and gradually to build up across the gap. This layer has a molten appearance in the way it changes shape under the influence of the discharge. Its growth continues until it extends about halfway across the gap when suddenly a shorting bridge appears as before.

Observations were made as functions of gap length and current of the average rates at which the anode growths bridged the gap. The data are plotted in Fig. 2. In each case the discharge was continued until the gap was bridged, or for five minutes if no bridge occurred. In the latter case the height of the growth was measured by microscope. The principal significance of these data is that they show the relationship of the bridging rate to position of the anode with respect to the negative glow: i.e., when the anode is immersed in the negative glow a rapid bridging occurs, but when it is outside the negative glow there is only a slow formation of a layer on the anode. Presumably it is the high electric field close to the cathode which causes the rapid bridging, by pulling some of the anode layer across by electrostatic attraction.

Anode growths have been observed also in Cu, Pd, Pt, Al, and W. The rates of growth for these metals are much less than for Ag, but they all bridge the gap after several minutes of discharge, with the exception of W. For W the discharge keeps moving to new spots on the anode and makes only a small growth at each spot.

The cathode area under the negative glow remains clean and bright while the anode process takes place. An exception, noted later, is when some of the anode material is detached in large chunks and is pulled across to the cathode. Since the cathode receives most of the energy of the discharge it becomes quite hot, and the area outside the negative glow darkens. The anode remains relatively cool.

CATHODE SPUTTERING

Three observations made us first suspect cathode sputtering as the cause of the growths. These observations are: (1) The build-ups are on the anode; (2) there is a slight cratering of the cathode when the discharge lasts several minutes; and (3) of the metals tested silver has the fastest bridging and is also the fastest sputterer.

The effect of sputtering on the fast bridging with silver was tested by making one electrode tungsten and the other silver. Güntherschulze[2] gives sputtering rates

[2] A. Güntherschulze, Z. Physik **36**, 563 (1926).

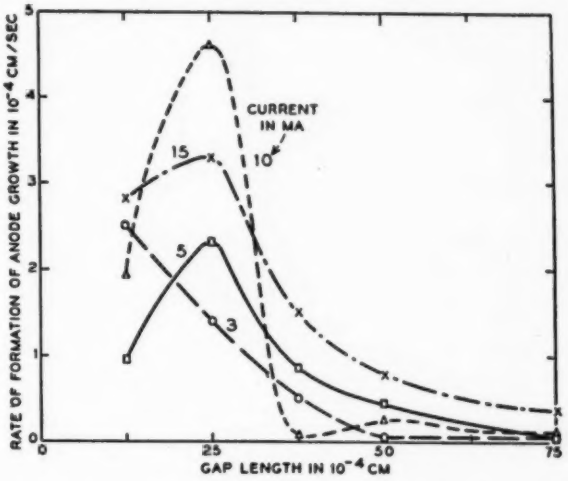


FIG. 2. Average rates of bridging caused by anode growths. Silver electrodes.

for a number of metals in the abnormal glow (small cathode area and high gap voltage). In the normal glow in the experiments reported here the sputtering rates are much lower, but, assuming that they keep about the same relative rates, the silver sputtering is expected to be an order of magnitude greater than that of tungsten. Therefore, with silver as the cathode the sputtering rate is high, and if the bridging is caused by sputtering we should then have the fast bridging. When the polarity is reversed with tungsten the cathode the sputtering rate is low and the bridging should be slow. This was tried in both air and nitrogen. Table I shows the results.

From this it is seen that in air the dependence of bridging rate upon polarity is just the opposite of what was expected from sputtering. In nitrogen however, the slow bridging when silver is the cathode, and the absence of bridging when tungsten is the cathode, are in agreement with the relative sputtering rates. Clearly then, although cathode sputtering occurs at a low rate, it is not the principal cause of the fast bridging. A chemical effect requiring the presence of oxygen is indicated, and an additional test was made of this conclusion by adding small amounts of oxygen to nitrogen (prepurified tank nitrogen of only 0.001 percent oxygen content). It was found that the addition of as small an amount as one part in a thousand increased the bridging

TABLE I. Sputtering test: 0.0025-cm gap, 0.010-amp current.

Electrodes −	Electrodes +	Sputtering rate (Güntherschulze) (μ g/coulomb)	Bridging in Air	Bridging in N_2
Ag	W	205 (high)	Slow 150–600 sec	Slow 270 sec
W	Ag	16 (low)	Fast 2–10 sec	None in 900 sec
Ag	Ag	205 (high)	Fast 2–10 sec	Slow 375–500 sec

TABLE II.

Compound	hkl planes	Reflecting power	Exper. intensity	Relative volumes
Ag	111	197	1.00	5.3
$AgNO_2$	110,020	31.4	0.30	10
Ag_2O	111	92.4	0.10	1.1

rate to eight times that in nitrogen alone. A similar experiment was tried with small amounts of nitrogen in oxygen, to see if nitrogen is required. In such an atmosphere the glow discharge could not be established, as a continuous sparking occurred, and hence no data were obtained. The question was answered by analysis of the anode growths as described in the next section.

DIFFRACTION ANALYSIS OF THE ANODE GROWTHS

Since the amount of material formed on one electrode was very small, for x-ray analysis material was collected from a number of extended runs. Using a Debye-Scherrer camera and FeKα radiation, patterns were obtained which proved to be of three substances: Ag, $AgNO_2$, and Ag_2O. From the known structures of these, their reflective powers for selected atomic planes were calculated for the type of camera used. By comparison of these with the observed intensities the relative volumes were estimated, with the results shown in Table II. It is to be noted that $AgNO_2$ is the most abundant material in the anode growths.

Both $AgNO_2$ and Ag_2O are reported to decompose at relatively low temperatures, and to obtain more information about this a quantity of each was prepared by chemical methods. Heating $AgNO_2$ at 70°C for 24 hours and then testing by x-ray diffraction showed that only a small percentage had decomposed, but at 110°C in one hour it had nearly all changed to silver. Ag_2O reacted similarly.

QUANTITATIVE MEASUREMENTS OF THE REACTIONS

A study of the electrode reactions was made from measurements of the changes produced in the weight of the electrodes. In the 25×10^{-4}-cm gap the bridges formed between silver electrodes were too small for such measurements, and hence wider gaps and longer discharge times were used, a 0.010-amp discharge for 20 minutes in a 100×10^{-4}-cm gap being common. The resulting changes were only a few micrograms, and it was felt that an accuracy of at least 1 microgram was necessary. No chemical balance of this accuracy was available to us, and consequently a torsion balance was constructed for the purpose. In it were combined several of the best features from the designs available in the literature, and using electrodes of about 5 milligrams in weight, changes in weight were measured to an accuracy of 0.5 microgram. The electrodes were again small rods, 0.076 cm in diameter, about 0.05 cm long for tungsten and 0.10 cm long for silver.

TABLE III. Electrode weight changes in the glow discharge (0.10-cm gap; 12-coulomb discharge).

Electrode material		Gain (+) or loss (−) in weight (micrograms) in nitrogen			in air		
Anode	Cathode		Anode	Cathode		Anode	Cathode
W	W	(1)	+0.1	+0.4	(5)	+ 2.4	0
Ag	W	(2)	+0.6	−0.6	(6)	+ 1.2	+ 5.7
W	Ag	(3)	+3.6	−4.5	(7)	+ 6.5	−45.6
Ag	Ag	(4)	+2.0	−5.5	(8)	+14.3	−13.7

Separation of anode and cathode reactions were made as before by using tungsten opposite silver, and chemical effects were separated from sputtering by comparing the results in air with those obtained in pure nitrogen. In Table III are summarized the changes in weight produced by discharges of 12 coulombs. In Fig. 3 is a graphical representation of the data.

Looking first at the results obtained in nitrogen it is seen from (1) and (2) that when the cathode is tungsten there is no appreciable sputtering loss, and both

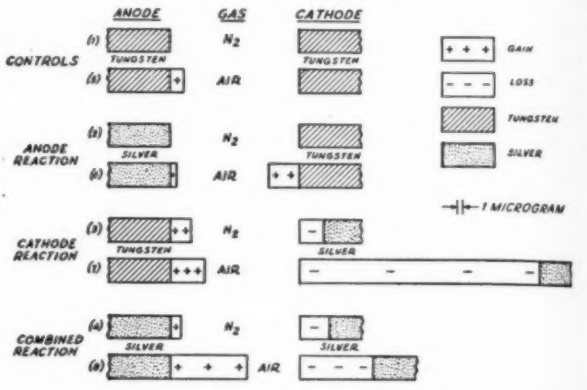


FIG. 3. Electrode weight changes resulting from glow discharges of 0.010 amp for 20 minutes. Electrode separation 100×10^{-4} cm. The electrodes are represented as opposite each other. Extensions toward the opposing electrodes are proportional to weight gains, and recessions away to weight losses. Changes in electrode separation are not represented.

tungsten and silver anodes have no appreciable reaction with the nitrogen. The changes listed are of the order of accuracy of the measurements. From (3) and (4) however, we see that when the cathodes are of silver, measurable sputtering occurs. The anode gains in these cases must be metal sputtered from the cathodes, since (1) and (2) show no anode reactions.

Turning now to the results in air it is seen from (5) that again there is no sputtering from tungsten cathodes, but that tungsten anodes make small but significant gains. A number of anodes were examined in the microscope and small mounds of yellowish white powder were found. These were probably WO_3, but a diffraction analysis was not made, as its identity is not sufficiently important. On changing to a silver anode as in (6) there are $AgNO_2$ anode growths. The cathode gain is explained by the observation that with the electrode separation used in these tests some of the material in these growths breaks off and jumps to the

cathodes instead of forming shorting bridges. This jumping was observed by microscope. With the electrodes reversed as in (7) there results a large loss from the silver cathode, about ten times as much as in nitrogen in (3). This indicates a chemical attack of the cathode in addition to the normal sputtering. Güntherschulze and Betz[3] have reported such a chemical effect. As to the 6.5-μg anode gain in (7), if the 2.4-μg gain of weight from oxide formation as found in (5) is subtracted, then only 4.1 μg, or 9 percent of the metal lost from the cathodes reaches the anodes. When both electrodes are silver in air, as in (8), there results the triple process of anode build up, loss from the cathode by chemical action, and loss from the cathode by sputtering. If one adds the separate anode and cathode results as obtained from (6) and (7) one expects the anode to gain about 5 μg and the cathode to lose about 40 μg, as against the actual results of 14.3 μg and 13.7 μg, respectively. The reason for this discrepancy is undetermined.

Turning again to (6) a minimum estimate can be made of the amount of $AgNO_2$ and Ag_2O which forms on the anode. The x-ray diffraction results give the composition of the growths as $AgNO_2$, Ag and Ag_2O in the ratios 10:5:1. From this they are about 80 percent Ag and 20 percent N and O. If it is now assumed that all of the material which leaves the anode is deposited on the cathode, and is there reduced to silver, then 7.0 μg must leave the anode to deposit 5.7 μg on the cathode. Since the anode shows a net increase of 1.2 μg then the total weight of material added to the anode is at least $7.0+1.2=8.2$ μg. Since this increase can come only from the addition of N and O from the air (as the cathode is tungsten there is no appreciable sputtering from it) the total anode growth must weigh at least 41 μg. This is at the rate of 3.4 μg per coulomb, since the discharge is 12 coulombs. This is a minimum estimate, because very likely not all of the material which leaves the anode reached the cathode, which if allowed for, would revise the estimate upward.

In (3) and (4) occurred the normal sputtering for Ag, at the rate of 5 μg for 12 coulombs, or about 0.4 μg per coulomb. Subtracting this from the cathode loss in (7) we have 41.5 as the loss by chemical attack, or 3.5 μg per coulomb.

[3] A. Güntherschulze and H. Betz, Z. Electrochem. **44**, 248–253 and 253–255 (1938).

DISCUSSION

The 0.4 μg/coulomb rate measured in nitrogen has been assumed to be pure sputtering. Comparison with the literature is difficult because of the lack of data for sputtering in the normal glow. The value of 205 μg/coulomb given by Güntherschulze[2] is for the *abnormal* glow at 850 volts. Rockwood[4] and Townes[5] give data on the sputtering of the active cathode surface in the normal glow discharge of cold cathode tubes. The values vary from about 0.0004 μg/coulomb for 0.010-amp current to 0.4 μg/coulomb for 0.155-amp current. The potential on those tubes was only 70 volts so that the value of 0.4 μg/coulomb obtained here for Ag in N_2 at 0.020 amp and 250 volts seems reasonable as pure sputtering.

It is very likely that the formation of the $AgNO_2$ on the anode involves some of the reactions occurring in the "fixation of nitrogen" process. In the electric arc the oxide NO is formed, which oxidizes further to NO_2 when cooled. In the glow discharge Holz and Müller[6] have identified both NO and NO_2 by means of band spectra. Brewer and Kueck[7] found in a glow discharge in an atmosphere of N_2+2O_2 at room temperature, that when equilibrium had been established there was 1.5 percent of NO_2 present. Presumably this oxide attacks the anode and forms $AgNO_2$. Then as the layer build up it is partially decomposed, and the result is a mixture of $AgNO_2$ and Ag. The small amount of Ag_2O in this layer may form independently.

The loss at the cathode ascribed to a chemical effect is several times that of pure sputtering. No direct evidence of the nature of this effect has been obtained. It is, however, altogether probable from the observations upon the anode that the same reaction to form $AgNO_2$ takes place at the cathode, but that the ion bombardment and the very considerable heat dissipated on the cathode (much more heat than on the anode) removes the nitrite as fast as it is formed.

During the course of this work the author received many valuable suggestions from L. H. Germer, to whom he expresses sincere thanks. He also wishes to thank G. E. Reitter for constructing the torsion balance.

[4] G. H. Rockwood, Trans. Am. Inst. Elec. Engrs. **60**, 901 (1941).
[5] Charles Hard Townes, Phys. Rev. **65**, 319 (1944).
[6] W. Holz and R. Müller, Ann. Physik **34**, 489–520 (1939).
[7] A. Keith Brewer and P. D. Kueck, J. Phys. Chem. **37**, 889–96 (1933).

New Booklets

Now available to interested readers is the latest booklet (DT-337) of the Fish-Schurman Corporation, describing *Steelset* diamond tools used for drilling holes and biscuit cutting and internal grinding of glass, porcelain, ceramics, quartz, etc. Address requests to Murray Yawitz, Fish-Schurman Corporation, 230 East 45 Street, New York 17, New York.

Reflecting Surface to Simulate an Infinite Conducting Plane

S. J. RAFF
U. S. Naval Ordnance Laboratory, Silver Spring, Maryland*
(Received November 2, 1950)

In measuring small microwave reflections back to the transmitting antenna, a finite reflecting surface which simulates the reflection of an infinite plane is desirable for calibration. Treating a flat reflecting surface by the methods of physical optics, the reflection can be represented as the sum of the contributions of Fresnel zones. Owing to the slow rate of decrease of reflections from successive zones, the shape of the surface has considerable effect on the accuracy with which its reflection simulates that of an infinite plane.

By means of the calculus of variations, an optimum shape is designed which, for a given antenna pattern, reflector size, and range of antenna to reflector distances, will give the closest simulation of the reflection of an infinite plane. Such an optimum surface of 23 wavelengths extreme dimensions has been constructed for use at 25 wavelengths from a dipole antenna. The maximum simulation error was calculated to be ±3 percent. The reflection from this surface has been measured and is within ±4 percent of that of an infinite plane over the range of 15 to 35 wavelengths from the antenna.

INTRODUCTION

WHILE procedures for testing the performance of microwave reflectors by means of the Naval Ordnance Laboratory electromagnetic model range were being developed, it became necessary to design, for calibration, a flat surface to simulate an infinite reflecting plane. The circuit of the model range operates on the wheatstone bridge principle, using a side outlet or "magic tee" as the basic bridge circuit. The same antenna is used as transmitter and receiver, and the reflections are measured by their effect on the antenna impedance and consequent effect on the balance of the tee. Since the model range uses a stationary antenna and moving target, the size of the surface to be used for calibration is limited. Some preliminary attempts were made to determine experimentally what size surface was required to give the same reflections as an infinite surface to an accuracy of ±3 percent, at about 25 wavelengths from the antenna. The results indicated that unless considerable attention were given to the shape of the surface, its size would have to be many times the distance from the antenna to target. This report contains the analysis required to determine the optimum shaped surface and the experimental results obtained in tests of two simple reflecting surfaces as well as the theoretically derived optimum one.

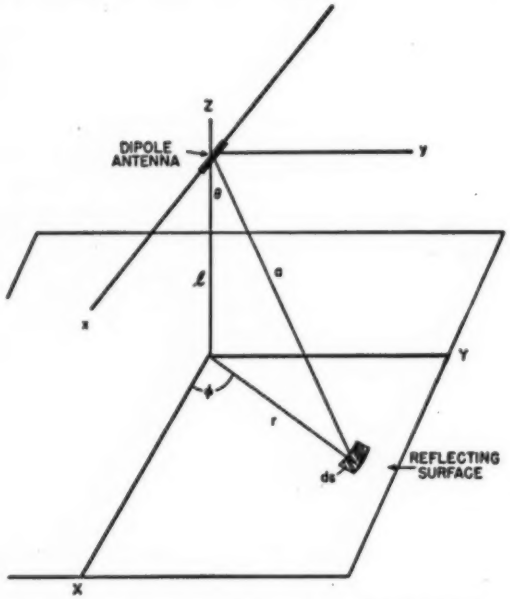


FIG. 1. Configuration of elementary reflecting area.

GENERAL METHODS OF CALCULATION

For present purposes we will be dealing with plane conducting surfaces whose dimensions and distances from the antenna will be of the order of 20 wavelengths. It is apparent that edge effects are negligible and that the wave front cannot be assumed plane over the surface. We are led, therefore, to the methods contained in the Fresnel diffraction theory. Referring to Fig. 1, the reflection from the shaded element of area can be expressed as†

$$dE = E_0 D(\mathfrak{a}) \Theta(\theta) P(\mathfrak{a}/\lambda) dS, \tag{1}$$

where

$dS = r d\phi dr$ is the area of the shaded element.

$D(\mathfrak{a}) = 1/\mathfrak{a}^2$ is the distance factor expressing the fact that the reflected field strength falls off as the inverse square of the distance from the reflecting element.

$P(\mathfrak{a}/\lambda) = \exp\{2\pi i[(2\mathfrak{a}/\lambda) + (1/2)]\}$ is the phase factor, where $2\mathfrak{a}/\lambda$ is the total number of wavelengths from the source to the reflecting element and back, and $1/2$ represents the change of phase on reflection from a good conductor.

E_0 represents the reflection from a unit area in the $\theta = 0$ direction from the source, at unit distance, and with a phase factor of unity.

$\Theta(\theta)$ is the obliquity factor expressing the dependence

* Submitted to the University of Maryland in partial fulfillment of the requirements for the degree of Master of Science in Physics.

† In Eq. (1) and the rest of the paper, the script $\mathfrak{a}$ refers to the a which appears in Fig. 1.

of the reflection from a unit area on the orientation of the position vector from the source.

All these factors, except $\Theta(\theta)$, and their values may be found in most elementary physical optics texts.[1] In most elementary diffraction problems the obliquity factor, $\Theta(\theta)$, may be assumed constant, since its variation over the area of integration is small compared with the variation of other factors. Here, where the distance from the source and the dimension of the reflector are of the same order of magnitude, the solution will be considerably influenced by the choice of the obliquity factor. This factor can be evaluated as the product of f_1 and f_2, where f_1 is the product of the voltage transmitting and receiving pattern of the antenna, or the power pattern; and f_2 is a similar pattern for the element of area: f_2 can be derived by the methods of Schelkunoff.[2] The element of area is assumed to be part of a large surface for purposes of calculating the current flow on it due to the incident wave, and the reflected wave is taken to be the field radiated by this current flow on the element of area. This procedure of calculation, for a dipole antenna in the configuration of Fig. 1, leads to the formula:

$$f_1 f_2 = 2\cos\theta\,(\cos^2\theta + \sin^2\phi\,\sin^2\theta).$$

In order to simplify the mathematics to follow it is necessary to assume that all surfaces to be treated here have circular symmetry. We can then take $\Theta(\theta)$ as the average $f_1 f_2$ over ϕ. This gives

$$\Theta(\theta) = \cos\theta + \cos^3\theta.$$

Substituting values for each of the factors in Eq. (1), evaluating θ and r in terms of l and $\mathfrak{a}$, and integrating over ϕ gives

$$dE = E_0(2\pi/\mathfrak{a})[(l/\mathfrak{a}) + (l^3/\mathfrak{a}^3)] \times \exp\{2\pi i[(2\mathfrak{a}/\lambda) + (\tfrac{1}{2})]\}d\mathfrak{a}. \quad (2)$$

Following the method of Fresnel and integrating over $\mathfrak{a}$ by half-period zones, we can express the reflection from a surface as

$$E = \sum_n \frac{E_0\lambda}{2i}\left(\frac{l}{\bar{a}_n^2} + \frac{l^3}{\bar{a}_n^4}\right)(-1)^n \exp\left[2\pi i\left(\frac{2l}{\lambda} + \frac{1}{2}\right)\right],$$

where each term in the summation is the reflection from a Fresnel zone, and $\bar{a}_n$ is the average value of $\mathfrak{a}$ over the zone,

$$\bar{a}_n = l + (2n-1)\lambda/8.$$

If the summation is extended to $n=\infty$, it can be readily shown that the sum is half the first term.[1] The summation from $n=1$ to ∞ is the reflection from an infinite plane. For $l \gg \lambda$ this reduces to

$$E_\infty = (E_0\lambda i/2l)\exp\{2\pi i[(2l/\lambda) + (\tfrac{1}{2})]\}.$$

[1] See, for example, F. A. Jenkins and H. E. White, *Introduction to Physical Optics* (McGraw-Hill Book Company, Inc., New York, 1937), p. 173.

[2] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), p. 358.

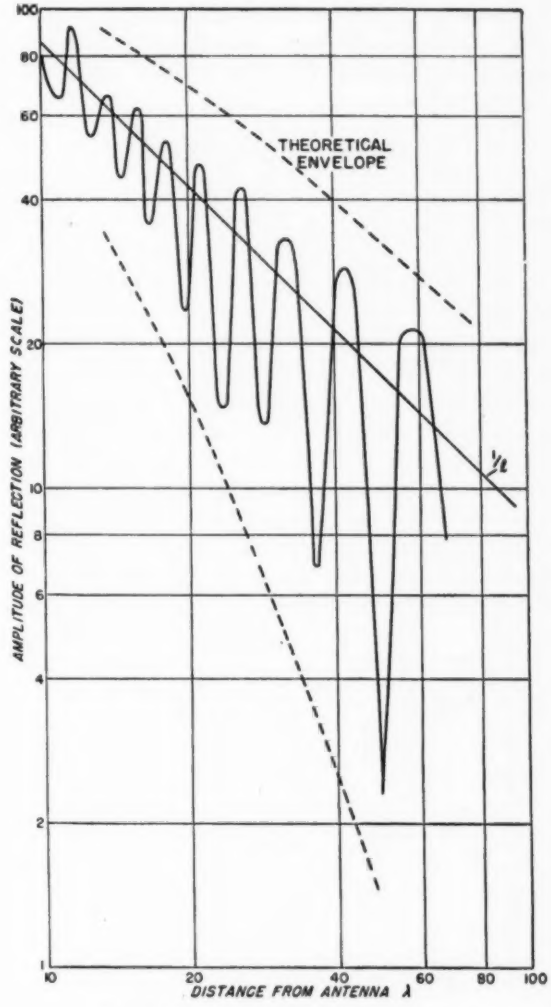


FIG. 2. Reflection from circular surface.

A circular disk is the only shape of surface which can contain a finite integral number of Fresnel zones. In this case the reflections can be found most simply by evaluating the difference between the circular surface reflection and that of an infinite plane. This difference represents the error involved when the circular surface is used to simulate an infinite plane. It is called δ and is simply the contribution to the reflection which would be made by the zones beyond the periphery of the surface. If z is the number of zones on the circular surface,

$$\delta = \sum_{n=z+1}^{n=\infty} \frac{E_0\lambda}{2i}\left(\frac{l}{\bar{a}_n^2} + \frac{l^3}{\bar{a}_n^4}\right)(-1)^n \exp\left[2\pi i\left(\frac{2l}{\lambda} + \frac{1}{2}\right)\right].$$

Since the summation is to $n=\infty$, the sum is, like the previous sum, half of its first term. Thus, the error or δ can be expressed in term of E_∞ as

$$\delta = \pm\tfrac{1}{2}E_\infty\left\{\frac{1}{[1+(2z+1)\lambda/8l]^2} + \frac{1}{[1+(2z+1)\lambda/8l]^4}\right\}. \quad (3)$$

This formula has been used to obtain the dotted lines in Fig. 2, which illustrate the calculated maximum error or δ as a function of l for a circular surface of 24λ diameter.

OPTIMUM SHAPED SURFACE

From the form of the error function derived in the preceeding paragraph it is apparent that the error would be less if the series representing the reflections from successive zones converged more rapidly. For example, some rough calculations have been made for square surfaces where the periphery, because of its shape, must intersect several zone boundaries, and so superimpose an artificial convergence on the series by eliminating successively larger portions of the last few zones. These results indicate that the δ of the square is considerably less than that of the circumscribed circular surface. This is borne out experimentally (see Fig. 3).

This idea of superimposed convergence can be carried further by calculating for a surface of given outer dimension and distance from the antenna, the optimum

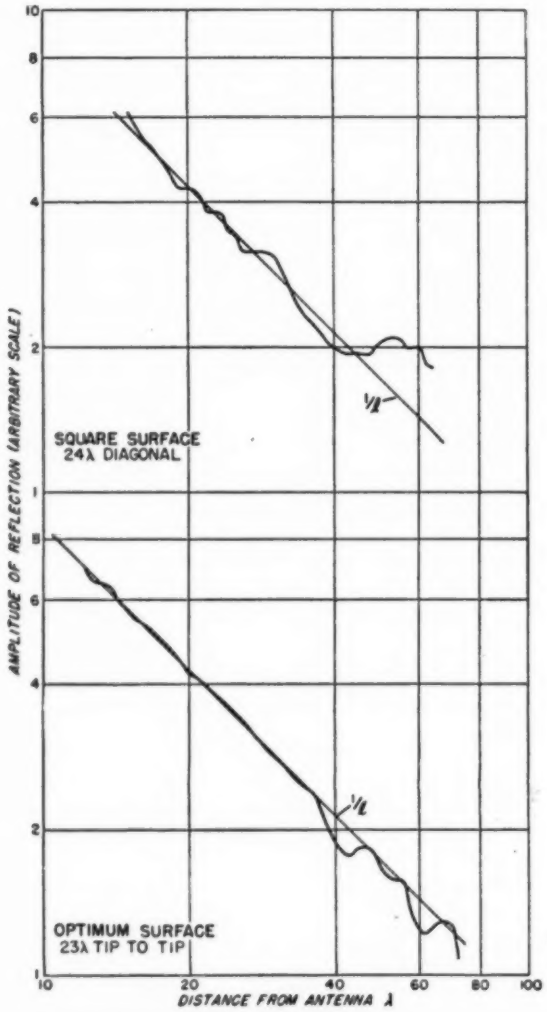


FIG. 3. Reflection from square and optimum surfaces.

convergence. For example, let $f(r)$ represent the fraction of the area in any ring $2\pi r dr$ which reflects; the remainder of the ring will be cut away in shaping the surface. For this purpose $f(r)$ must have the following properties: it must be unity at $r=0$ and zero at $r=r_0$, where r_0 is the radius of the circle circumscribing the figure. Also, if the second member of Eq. (2) is multiplied by $f(r)$ and the integration performed, the difference between this integral and the reflection from an infinite plane must be a minimum. Using $f(\alpha, l)$ instead of $f(r)$ permits us to perform a preliminary integration by parts which gives

$$E=E_\infty-\int_l^{\alpha_0}\frac{d}{d\alpha}\left[E_0\frac{\lambda}{4i\alpha}\left(\frac{l}{\alpha}+\frac{l^3}{\alpha^3}\right)f(\alpha,l)\right]\times\exp\left[2\pi i\left(\frac{2\alpha}{\lambda}+\frac{1}{2}\right)\right]d\alpha. \quad (4)$$

It is thus clear that for best simulation of infinite plane reflection the remaining integral should be a minimum subject to the stated end conditions on $f(\alpha, l)$. However, it should be noted that, although l is constant in the preceeding integration, a surface is desired which will reproduce an infinite plane over a range of l. This affects considerably the selection of $f(\alpha, l)$. For example, one could select an abrupt function which would make the integral zero; but on a change of l, the configuration of the surface would remain the same and $f(\alpha, l)$ will change, or stated differently, the Fresnel zones will move across the surface.

Thus, for purposes of minimizing, we should consider the exponential factor and the remainder of the product as though they were functions of different independent variables, which require separate minimizing of their magnitudes. Of course, we have no control over the exponential factor; and so the minimum condition which is sought is that the integral of the magnitude of

$$(d/d\alpha)\{(E_0\lambda/4i\alpha)[(l/\alpha)+(l^3/\alpha^3)]f(\alpha,l)\}$$

be a minimum. Or that

$$\int_l^{\alpha_0}\left\{\frac{d}{d\alpha}\left[\frac{E_0\lambda}{4i\alpha}\left(\frac{l}{\alpha}+\frac{l^3}{\alpha^3}\right)f(\alpha,l)\right]\right\}^2 d\alpha$$

be a minimum subject to the conditions that $f(l, l)$ is unity and $f(a_0, l)$ is zero. This is a problem in the calculus of variations which can be restated in a simpler form by letting

$$(E_0\lambda/4i\alpha)[(l/\alpha)+(l^3/\alpha^3)]f(\alpha,l)=y.$$

The application of Euler's equation, integration, and evaluation of constants gives

$$f(\alpha,l)=2(\alpha-a_0)\alpha/l(l-a_0)[(l/\alpha)+(l^3/\alpha^3)].$$

EVALUATION OF ERROR

We can evaluate the error, or δ, when such an optimum surface is used by substituting the above value

of $f(\alpha, l)$ in the second term of Eq. (4) and evaluating that term. Substitution gives

$$\delta = -\int_{l}^{a_0} \frac{E_0 \lambda}{2il(l-a_0)} \exp\left[2\pi i\left(\frac{2\alpha}{\lambda}+\frac{1}{2}\right)\right] d\alpha.$$

After E_∞ is factored out, this integrates to

$$\delta = [\lambda E_\infty / 4\pi i(a_0-l)]\{1-\exp[4\pi i(a_0-l)/\lambda]\}.$$

Because of the factor i in the denominator, the real part of the term in braces indicates a component of δ in quadrature with E_∞. Since, in general, $\lambda/4\pi(a_0-l) \ll 1$, $E_\infty \gg \delta$; and this quadrature component of δ has little effect on $|E_\infty+\delta|$, which is the magnitude of the reflection from the surface. The observable error depends on the imaginary part of the term in braces, which has the maximum value 1. Therefore, the maximum observable error will be

$$\delta_{max} = \pm E_\infty / z\pi,$$

where z is the number of Fresnel zones represented on the surface; i.e.,

$$z = 4(a_0-l)\lambda.$$

DESIGN CONSIDERATIONS

For model range calibration a limiting accuracy of 3 percent was considered satisfactory, and a range of calibration distances was chosen to have its design center at $l = 25\lambda$. The maximum error of 3 percent gives the number of zones equal to 10 and

$$f(\alpha, l) = 2(\alpha^2 - 27.5\alpha)/2.5[(25/\alpha)+(25^3/\alpha^3)]25;$$

$f(\alpha, l)$, as defined when introduced, is the fraction of the area of any ring $2\pi r dr$ which reflects, expressed as a function of $a = (l^2+r^2)^{\frac{1}{2}}$. The remaining area of the ring will be cut away. In view of the assumption of circular symmetry made in deriving $\Theta(\theta)$, this area should properly be removed so as to retain circular symmetry. However, it has been demonstrated experimentally that symmetry about the $\phi = 0°$, 45°, 90°, 135°, etc., axes is sufficient for the accuracy desired. Therefore, the unwanted area of the reflector was removed in four sectors, leaving the four-pointed shape shown in Fig. 4.

DISCUSSION

Figure 3 shows the result of measurements made on the derived optimum surface; Figs. 2 and 3 show the results for a circular and square surface of about the same outer dimension. The error involved in using one of these surfaces to simulate an infinite surface is represented by the deviation of the curve from the straight line $1/l$ drawn through it. The qualitative agreement with the theory is apparent. The circular surface gives the greatest deviation or δ. The square surface gives considerably less deviation than the circle; and the optimum surface, designed to give a maximum deviation of 3 percent in the vicinity of $l = 25\lambda$, gives an error of less than 4 percent between 15 and 35λ.

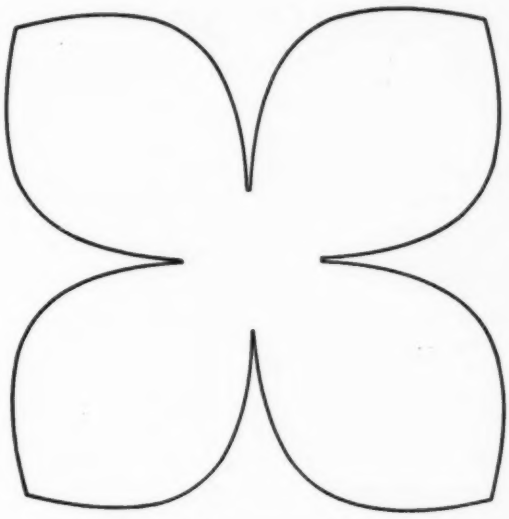


FIG. 4. Outline of optimum reflecting surface.

The experimental results on the circular surface are readily subject to a more accurate comparison with the theory. The maximum values of the error function δ have been calculated from Eq. (3) for this surface as a function of the distance from the antenna. The results of these calculations are indicated by the dotted lines on Fig. 2. If theory and experiment agreed perfectly, these dotted lines would be the envelope of the curve for the circular surface. The difference is probably due to the misalignment of the surface during measurement. The exact effect of this misalignment is difficult to calculate. However, the maximum error, indicated by the dotted lines in Fig. 2, can occur only when a zone boundary coincides with the boundary of the surface. Misalignment prevents this coincidence. The probable experimental error in alignment is 1/12 of one degree. That this order of magnitude misalignment can cause a significant variation in the experimental results has been demonstrated experimentally by making similar measurements with a misalignment of $\frac{1}{3}$ of one degree. The values of δ so obtained were generally about $\frac{1}{3}$ those shown in Fig. 2. It can be demonstrated both theoretically and experimentally that the alignment of the optimum shaped surface is far less critical.

Negative Coaxial Cylindrical Corona Discharges in Pure N_2, O_2, and Mixtures Thereof*

CHARLES G. MILLER† AND LEONARD B. LOEB‡
University of California, Berkeley, California

(Received November 2, 1950)

Direct-current corona studies were made on coaxial cylinders with inner cylinder at negative potentials in pure N_2, pure O_2, and mixtures of 1 percent 10 percent O_2 in N_2, and on clean dry air from 27-mm to 760-mm pressure. Observations of current potential relations from field intensified ionization currents to as near breakdown as possible were made noting thresholds for current transitions, pulsed discharges and other oscillographic as well as visual appearances. Negative wire corona does not exist as such in truly pure N_2. Unless a high series resistance was in the line once the critical region of potential was reached, the gap broke down to an arc. With current limiting resistor bombardment of the filament with positive ions from the low order pre-discharge released enough gaseous impurity so that after six minutes the resistor could be removed and the internal resistance of the negative ion space charge near the anode held the current at 2000 μa. The behavior indicated that breakdown was preceded by a low current Townsend discharge that cleaned up or conditioned the filament. In all the negative wire studies the onset thresholds vary with the past history of the wire, and the threshold potential at higher pressures overshoots the minimum operating potential of the discharge once it has set in. Studies with 1 percent O_2 indicate that the threshold discharge with clean gas is a continuous current with a diffuse glow which in time contracts to a cathode spot with Trichel pulses. Similar behavior applies to air at lower pressure. The negative wire in O_2 shows very heavy predischarge currents with relatively small current jumps when random Trichel pulses appear, and further small current jumps when the fixed spot with regular Trichel pulses appears. There is evidence for a continuous heavy background of Townsend discharge along the whole wire until the visible spot and heavy pulsed discharge appears. The duration of Trichel pulses in pure O_2 is two to ten times those in air with point to plane geometry and two to three times those in air with coaxial cylindrical geometry. Lower pressures in pure O_2 show many current transitions and transitions between discharge forms.

1. INTRODUCTION

THE apparatus developed for the study is described in a preceding article.[1] In that article there are also presented a set of definitions of terms that will be used freely in what follows. The reader is referred to the earlier article on positive coronas both for the description of the apparatus and the definition of terms.

The current data observed in initially pure N_2 as a function of voltage at various pressures cannot properly be represented by a series of curves as is the case for all the other gaseous combinations. The reason for this lies in the inherent instability of all cathode phenomena as Loeb has shown in detail elsewhere.[2,3] Thus, unless negative ion formation, electron space charge, or external limiting resistors intervene[2] a cathode, once it starts a self-sustaining Townsend discharge, runs progressively into a complete breakdown or arc over.

In general, the phenomena connected with such breakdown have not been observed because with dirty gases and/or adequate gap lengths the ion formation or electron space charges have limited the current. This was true for Weissler's[4] work in N_2 and H_2 with point-to-plane corona. In the present work with coaxial cylinders it was found impossible to make any studies of the gas *without limiting series* resistors or by soiling the gas. With clean wire and pure N_2 at 617 mm it was repeatedly observed that as the applied potential was gradually raised to around 4100 volts with no limiting resistor, a heavy discharge occurred, melting either the wire or a 10-ma protecting fuse. In contrast, at 750 mm with 4100 volts and 200 μa, a 3.5-megohm limiting series resistor permitted observation of a threshold with the appearance of a spot. There was no preliminary glow comparable to that observed by Weissler with a point-to-plane gap or as will later be described.[4] Thus, a study of the negative point with its unstable cathode must be made with a limiting resistor. For the negative current-voltage characteristic of such an unstable system can show controllable data only by putting enough series resistance into the circuit to give a net positive resistance, that is, by externally limiting the current. To get the true current-potential curves, the iR drop through the current limiting resistor may be deducted from the applied potential. This procedure is permissible and leads to reasonable values only *above the threshold*. There the current can be measured, is steady, and is not altering with time as is the case as the breakdown potential is approached from below for the first time in clean gas. Thus, no threshold currents were recorded in pure N_2, nor could the exact potential causing breakdown be established.§[5] Presumably, the currents were of the order of 10^{-13} ampere up to near the point where the Townsend discharge set in.

The character of the difficulties encountered is indicated by the events of Table I. Thus, at 607-mm pressure with 10.7-megohm limiting resistor, breakdown threshold occurred at 3565 volts. The post threshold current was 140 μa, requiring a sustaining potential *across the gap* of 2065 volts. The offset

* This work was sponsored by the Research Corporation of New York and in part by the ONR.

† Currently at the University of California at Santa Barbara.

‡ The experimental work was entirely done by Dr. Miller. Owing to the complex nature of many of the phenomena, much of the interpretation fell to the lot of Professor Loeb.

[1] C. G. Miller and L. B. Loeb, J. Appl. Physics **22**, 494 (1951).
[2] L. B. Loeb, Phys. Rev. **76**, 255 (1949).
[3] L. B. Loeb, J. Appl. Phys. **19**, 882 (1948).
[4] G. L. Weissler, Phys. Rev. **63**, 96 (1943).

§ Analogous results have recently been observed with pure N_2 in uniform fields by G. A. Kachickas and L. H. Fisher (reference 4).

[5] L. H. Fisher and B. Bederson, Phys. Rev. **75**, 1615 (1949); **78**, 331 (1950). Also, G. A. Kachickas, Phys. Rev. **79**, 232 (1950); B. Gaenger, Arch. Elektrotech. **39**, 508 (1949).

potential was 2000 volts at a current of 80 μa, indicating an initial overshoot in potential. In each case after the first run, which slightly soiled the gas, the discharge began as did Weissler's. That is, at threshold a faint diffuse glow down the length of the wire appeared which in time contracted to a single bright spot. The higher the potential above threshold, the shorter the time required for contraction to a spot. The contraction was not accompanied by a great change in current. This was doubtless owing to the larger limiting resistor used here and to the larger electrodes, since the large wire area was capable of giving a much greater total Townsend current than Weissler's point.[4] The first run with very pure N_2, on the contrary, did *not* show the diffuse glow but went right over to breakdown with spot. The currents increased to around 2000 μa with a 0.1-megohm limiting resistor. A 24-hour wait and outgassing the filament gave onset at 3100 volts and 3000 μa, and after 24 hours more it came on at 3278 volts and 1900 μa. Here again, the discharge contracted from an extended glow to the forms indicated in Fig. 1. In this case, it was clear that running the discharge on the wire for some time with limiting resistor, so that no breakdown occurred, eventually altered the system. Thus, when the resistor was removed, the internal space charge resistance had increased to a point where no arc over occurred. The discharge was steady, and at no time were any pulses observed. To check the conclusions, pure gas was introduced at 609 mm under conditions which had previously led to a spark with no limiting resistor. A discharge was run with 3.5-megohm resistor for 19 min at 800 μa. When the resistor was removed, the discharge set in at 3880 volts and 1600 μa as a glow that quickly contracted to a spot. That the changes caused come from the gas and not from changed conditions on the wire was shown by the fact that standing after a discharge had ceased did not restore the system to its original condition. Had the breakdown limiting action been caused by a gas film, this would have reformed in a few hours. Such action was observed where the positive wire and negative Ni cylinder recovered their high work function in a few minutes' standing after the cleansing discharge ceased. Glowing the wire as well as electron bombardment of the wire did not change it. Leaving the wire as it was, pumping out the contaminated gas and filling with pure N_2 again led to arc over unless a protective resistance was used.

It is, thus, clear that operating the discharge liberated gases like H_2O and O_2 from the central wire by bombardment with positive ions of high energies. In consequence, the gas in the chamber was sufficiently contaminated to give enough negative ions in the low field region to limit the current to about 2000 μa. The contamination did not produce enough negative ions to cause the discharge to fluctuate or show extinction as with Trichel pulses. The fouling was enough to cause the initial breakdown discharge to begin as a faint general diffuse glow down the whole wire. Presumably, this impurity acts on the wire to give a coating of higher work function which requires an initial Townsend discharge of low current density to clean it up yielding a breakdown threshold.

Fouling also assists in spreading the discharge down the wire by photo-ionization in the soiled gas. If N_2 is very pure, the photons spread the length of the cylinder

TABLE I. Purified nitrogen at 607-mm pressure effect of repeated discharges on threshold conditions.

I Limiting resistor (megohms)	II Onset voltage	III Post-onset current (μa)	IV Sustaining voltage for currents given in III	V Offset current (μa)
10.7	3565	140	2065	80
6.0	3600	200	2400	80
3.5	3700	320	2580	80
2.5	3650	410	2625	80
1.5	3650	600	2750	80
0.90	3600	800	2880	97
0.45	3550	1300	2965	90
0.29	3575	1600	3110	90
0.10	3438	2000	3238	100
0	3200	2300	3200	100
0[a]	3100	3000	3100	
0[b]	3278	1900	3278	
0[c]	3778	300		
0[d]	3180	1600		

[a] Wait 24 hours, glow out filament.
[b] Wait 24 hours more.
[c] Refill to check with purified gas above.
[d] Run 19 minutes at 800.

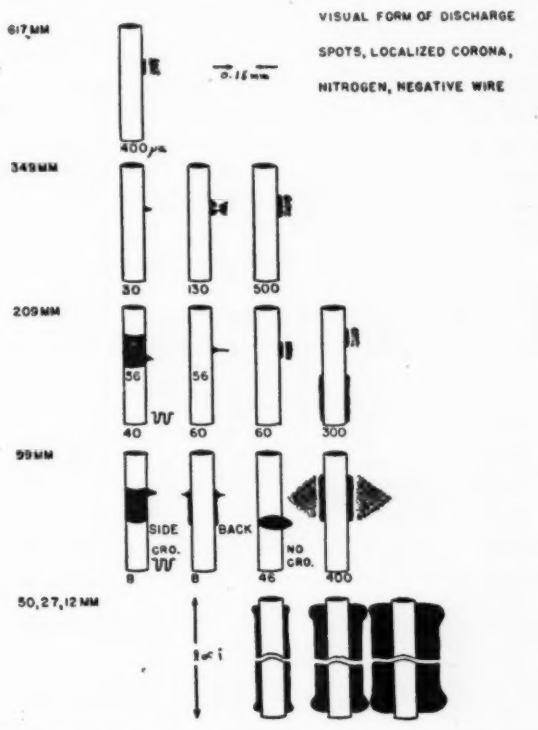


FIG. 1. Typical negative wire corona forms in pure nitrogen at various pressures and with various currents in microamperes indicated below the wire. These forms were all obtained with limiting resistors or else sufficiently soiled gas to prevent complete breakdown.

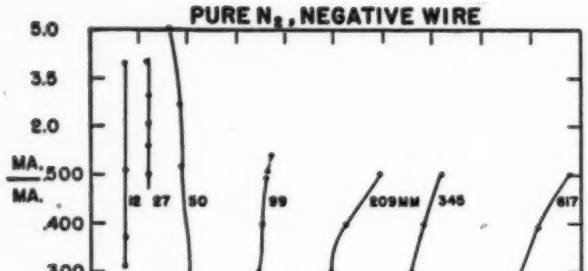


FIG. 2. Current-potential curves for pure nitrogen with a negative wire and limiting resistor. Different limiting resistors were used for parts of broken curves. Negative slope is of no significance and is caused by inaccuracies in computing IR drops in resistors.

but trigger no diffuse breakdown, as they produce no photo-electrons near the wire *except* right at the point where the avalanche took place. This follows from adverse geometry, fine cathode wire large cylindrical anode. Hence, when breakdown does occur in pure N_2, it starts at the place where the first intense avalanches and Townsend discharge began. The spot appears at the place where this discharge cleaned up the wire. This accounts for the low value of the Townsend discharge current with pure N_2 preceding breakdown to an arc.

If the gas is slightly soiled, the diffuse Townsend discharge current lies along the wire and is larger. It continues and the wire cleans up in a spot. This overvolts the spot, which increases emission, causing a concentrated discharge. In consequence, the general glow contracts to a spot. As to why high current density discharges of this sort remain in spots instead of spreading down the tube, the reader is referred to a recent paper of Loeb.[3]

While the negative wire discharge soils the gas, the positive wire discharge causes little contamination of the

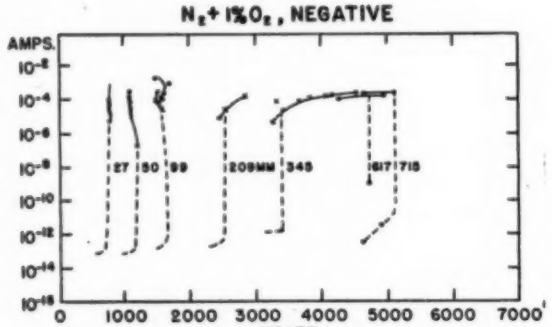


FIG. 3. Current potential diagrams for negative wire with one percent oxygen. Note the relatively steep rise of current below threshold. Note the heavy overshoot at higher pressures and the steep current voltage curves at low pressures where limiting resistors were required to prevent arc over.

gas. The Ni cathode about the wire has a low field of 230 volts/cm at its surface instead of 10^5 volts/cm as with the negative wire in this work. Low energy positive ions do not liberate gas from the metal.

In the relatively pure N_2 used, even with the contamination alluded to, the clean-up occurs at the threshold potential for photoelectric conditioned Townsend discharge if given time. In gases like N_2 with 1 percent O_2, the active impurities, e.g., O_2, NO_2, reform the coated surface at such a rate that there must be considerable current *density* at the spot to give and maintain a clean-up. Thus, the threshold for the large current jump and spot formation occurs at values of potential well above the Townsend threshold in such gases, while coinciding with the Townsend threshold in slightly soiled N_2.

From the foregoing it will be seen that the current-potential curves for the negative wire in pure N_2 or slightly soiled N_2 shown in Fig. 2 are not too significant. Here, the potentials are the applied potentials less the iR drops in the current limiting resistors. No importance should be ascribed to the slightly negative slopes of some of the curves.

The various visual appearances of the negative coronas under different pressures and currents are shown in Fig. 1. It is not profitable to speculate about these at any length. The diffusion broadening between 12 and 50 mm should be noted, as should the dimensions for spots, which in general correspond to those observed in negative corona from points, *viz.*, from 0.1–0.2 mm diameter. At higher pressures the Crooke's dark space is hinted at by the shapes, while the negative glow, faraday dark space, and positive column are more clearly delineated. The glow appears to spread around the wire with increasing currents. With the overloading of spots for very heavy currents, the appearance of multiple spots is manifest just as with negative point corona. With multiple spots the current is likely to lose its steady character and fluctuations in intensity are observed on the oscilloscope. The fluctuations are, however, of large amplitude and long duration, whereas Trichel pulses are of 0.4 or perhaps a few microseconds' duration. In any case, where steady currents are very large, what are small percentage superposed fluctuations can appear on the oscilloscope screen as large amplitude fluctuations, since the dc component is suppressed. In some cases, as at 209 mm, the discharge in a certain current range seemed to change back and forth from general glow to concentrated spot producing long period oscillations (about 0.01 sec) on the oscilloscope. Such phenomena are probably circuit conditioned.

Below 50 mm the intensity of positive ion bombardment of the wire owing to diffusion does not produce a spot, so that at threshold a regular glow takes its place. As the current increases, the glow spreads. The failure to give a spot results from the fact that the positive ion space charge and current density is no longer great enough to concentrate the discharge and maintain a

clean spot. Thus, the initial Townsend discharge in intensified form sustains itself as a glow discharge at higher potentials.

2. NEGATIVE WIRE N_2+1 PERCENT O_2

It is to be expected that with 1 percent contamination of O_2 the situation observed in slightly soiled N_2 will be enhanced. Thus, no complete breakdown will be expected in the absence of limiting resistor. Enough negative ion formation might be expected to cause interruption or pulsing of the discharge, particularly if concentrated in points. Diffuse discharge along the wire could hardly produce enough space charge density of negative ions to choke itself off. With this consideration in mind, it is possible to refer to the current potential curves of Fig. 3, which in part tell the story. Unfortunately, these measurements were made near the end of the series, and trouble with the electrometer which required major repair precluded taking too much data in the low current region. It appears, however, that the

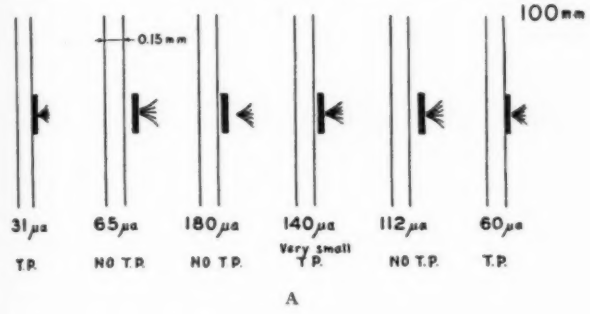


FIG. 4. A and B show characteristic structures generally observed with negative points.

currents rose relatively steeply from 10^{-13} ampere, indicating the presence of considerable photoelectrically conditioned Townsend discharge below threshold. The threshold was, however, complex. When the potential was applied and raised after 30 or more minutes of standing, the discharge appeared at a suitable potential which was much larger than that for pure N_2 and almost equal to that for air. The discharge appeared as a stationary spot glow and showed no oscilloscope pattern indicating a steady current. The current was fairly high. At 715 mm the threshold appeared at 5150 volts with a rise from 10^{-10} ampere to 100 μa. Figures 4A and B show the characteristic structure generally observed with negative points, the faraday dark space being less prominent because of the geometry and method of viewing. After this discharge had run for half a minute, or when the potential had been reduced or the tube had been restarted *shortly* after the discharge extinguished, a new phenomenon appeared. Onset of the *steady corona* with a visible spot was *preceded* by an oscilloscopic pattern consisting of single pulses shown in Fig. 5 of duration of 7×10^{-4} sec randomly spaced. Radioactive triggering increased the number but did not alter the size. These pulses occurred over a range of 200 volts leading to *onset* of steady corona. Increasing potential led to visual corona at 20 μa; the oscilloscope pattern then showed regular pulses such as those of Fig. 6. These

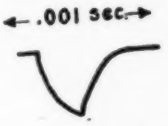


FIG. 5. Characteristic negative pulses observed just before steady corona onset with a single spot.

varied in frequency according to the following schedule:

at 10 μa	500 pulses/sec
20 μa	1300 pulses/sec
40 μa	2500 pulses/sec
80 μa	6400 pulses/sec
90 μa	9400 pulses/sec.

Beyond this further increase in potential caused the pulses to cease. It is believed these were Trichel pulses of long duration required in the presence of little O_2 to form enough negative ions in the low field region to choke off the discharge. The frequency depended only on the potential, or current, and was not affected by the limiting resistor. Steady running of the discharge for six minutes reduced the frequency at a given current presumably by producing substances reducing ion mobilities or increasing attachment. At 9000 cycles and 100 μa the pulses merged, giving steady corona, which is to be expected when the duration of the pulse is comparable with the time between pulses. However, this criterion cannot always be applied, since multiple spot sometimes occurs before one spot can be overloaded to the point of giving no pulses.

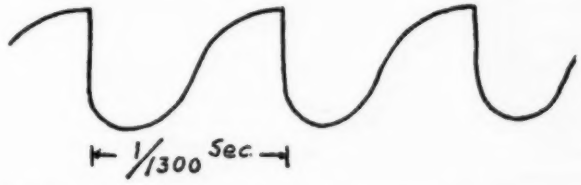


FIG. 6. Oscilloscope pattern with visual corona at 20 microamperes in N_2 plus one percent O_2.

With this description of the nature of the phenomena at the onset of heavy currents, it must next be noted from Fig. 3 that once the current is going, in *either mode*, reduction in potential carried the discharge at fairly high current down by 500 to 1000 volts before the discharge ceased. The *overshoot* was observed down to pressures of 200 mm. There was still some indication of it at 99 mm.

After the discharge had run for some time with fairly heavy current, apparently some electron attaching impurity was produced by chemical action in the gas. In

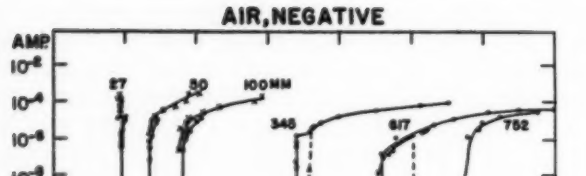


FIG. 7. Current-potential curves for negative wire in air. Note the very varied levels of initial prethreshold currents and the sharp and irregular thresholds with much overshoot extending down even to lower pressures.

consequence, enough ion formation with the low concentrations of O_2 took place to choke off the current after some 10^{-3} sec and gave Trichel pulses of long duration. If the gas stood for *30 minutes* or more after the first discharge building up the Trichel pulse forming impurities, these disappeared. Thus, the discharge started anew as a nonpulsing discharge. At the same time, in a *very few minutes* after the discharge ceased, the *conditioned spot became contaminated* and considerable overshoot of potential was needed again to clean up a spot and start the discharge. In fact, even with 1 percent O_2 considerable sustained current density is needed to form and keep a spot clean and active. The minimum pulse repeat rate of 10^3 per second is consistent with ion mobilities.

The change of the curves of Fig. 3 as pressure decreases is to be noted. Below 200 mm the slope of the current-voltage curves at threshold became so steep that the discharge could only be studied with a limiting resistor. This merely indicates that below 99 mm pressure with 1 percent O_2, except close to the threshold, the negative ion formation is not adequate to give the current limiting internal space charge resistance. Thus, the discharge needs external resistance to prevent arc over. At 99 mm without limiting resistor, threshold accompanied by Trichel pulses began at 1550 volts and some tens of microamperes. The current then rapidly rose to higher values, and Trichel pulses ceased. With limiting resistor the current could be followed from more than 100 μa with no pulses to down below 100 μa with Trichel pulses. The spot extinguished at 10–20 μa. Below 99 mm the discharge, once started, rose automatically to such magnitude as to blow a 10-ma fuse unless limiting resistors were present.

3. NEGATIVE WIRE IN AIR

As with N_2 and 1 percent O_2, the current potential curves in air with negative wire were, if anything, more irregular and confusing as seen in Fig. 7. The curve for 752 mm was made for a new filament and clean air. After it had been run, the *filament and gas had altered* so that the ensuing curves at 617 mm and below essentially represent an altered system. Another curve for a new filament and fresh air at 752 is shown in Fig. 8, where curves were taken with and without radioactive triggering. Currents were fairly constant at about 10^{-10} ampere from 3000 volts to near onset at 5700 volts. Radioactive triggering appeared to increase the currents by a factor of 2 except near threshold. There, radioactivity appeared to initiate Trichel pulses with a current increase somewhat below the values in the absence of radioactivity. Once the discharge had been initiated, the sequence of events was much more like those shown for 617-mm pressure and below. The currents appeared to be of the order of 10^{-12} amp and led to discharges showing much overshoot. In fact, the starting conditions varied from day to day depending on the preceding treatment and how long the wire had remained standing after discharge. The dashed rises of the curves at 617 mm and 345 mm indicate the complete absence of reproducibility. At about 10^{-12} amp and near 4400 volts an increase in potential would result in sudden sharp transition to currents of 0.1 to 1 μa. The transitions did not occur at once but only after waiting some time at the potential after the potential was raised above 4400. The time of transition was shorter the higher the potential above 4400 volts. Accurate values were difficult to fix and varied from time to time. At these thresholds the coronas showed continuous current with a slight ripple. The visual manifestation was a glow along the whole length of the wire, particularly if the system had been standing overnight. After some time the glow contracted to a spot, frequently with increase in current at the same potential. Spot appearance was marked by the appearance of Trichel pulses. If the discharge was interrupted for a short time only, it started at once as a spot at the same location and showed Trichel pulses. After some 10 or 12 hours of standing it began as a glow and ultimately contracted. Once the spot and Trichel pulses had formed, the potential could be reduced materially without much reduction in current until potentials of the order of 4400 volts were

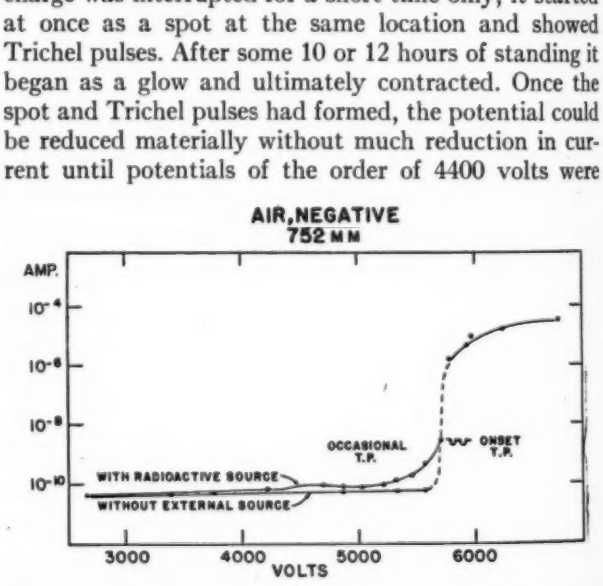


FIG. 8. Current-potential curves for negative wire corona in air, fresh gas, and wire near atmospheric pressure. Note the difference in approach currents below threshold caused by radioactive triggering.

reached. At this point the discharge extinguished and currents went to low values. Similar situations applied at 345 and presumably at 200 mm. If the discharge began as a glow and the potential was reduced *before* the spot formed, the extinction curve of the glow paralleled the Trichel pulse extinction curve though with lower current values. The point occupied by Trichel pulse glows after cutoff showed the characteristic craters with bright central spot and dusty oxide rims observed by Peek[6] and by Loeb, Kip, Hudson, and Bennett in point to plane corona.[7]

The observations with clean wire and fresh air by their nature indicate that either a field intensified current or a low value Townsend current governed by Eq. (2) of reference 1 is active. The effect of heavy radioactive triggering should not increase a self-sustaining Townsend discharge current, while it should increase a field intensified current by a large factor. The low increase of current by triggering thus indicates that a Townsend discharge was acting, but that it tended to be interrupted or extinguished rather readily. If such a discharge is interrupted or does not spread readily along the wire in virtue of a heavy absorption of photons, then triggering would increase the current. The earlier initiation of pulses by triggering also seems to point some sort of interruptions in the Townsend discharge whose reignition is facilitated by new triggering electrons at different spots. Such interpretation appears plausible because of the obvious rapid aging of the wire once a heavy discharge has been run. Possibly the 10^{-10} amp Townsend discharge is interrupted by deactivating gaseous agencies produced by it. Thus, the wire and, perhaps, even the air are so modified that the predischarge current is reduced by a factor of 100 when the heavier discharge has acted for some time. Whether the lower currents of 10^{-12} ampere are Townsend discharges with a badly oxygen-contaminated surface or just field intensified currents, one cannot say. They extend uniformly over the length of the wire as indicated by the current distribution to the guard rings. Their rise with increasing potential in any case is very slow, appearing to indicate a possible equilibrium being achieved between deactivating the surface by chemical discharge products and cleaning it up by ion bombardment. Apparently, at or near 4400 volts at 617 mm, the clean-up action begins to exceed the contaminating activity. Hence, in time the surface cleans up sufficiently to give rise to a threshold for a diffuse visible discharge with a very large current jump and overshoot. At higher potentials this clean-up occurs more rapidly. Thus, with limitations of human patience, it becomes difficult to fix a threshold exactly, especially as this varies from day to day. As would be expected, with the heavier diffuse current densities above threshold it is not surprising that this current regime is unstable. For accidental local fluctuations in current density will lead to the final conditioning of a spot with high overvolting and a Trichel pulse regime. This, being heavily overvolted, suffices to carry the current at lower potential; and the glow extinguishes, yielding a spot. The interpretation given appears to be substantiated by the various properties of the discharge, as indicated by the observations cited, such as aging, overshoot, offsets, etc. In principle, the nature of the mechanisms is not very different from those in N_2 with 1 percent O_2. Trichel pulses are more readily formed and more rapidly choke off the discharge in air. Again, with more O_2 the surfaces are more rapidly and completely contaminated and require greater currents in air for the ratio of cleansing impacts by assumedly N_2^+ ion impact is now counteracted by the more frequent deactivating action of the O_2^+ and O^+ ions. Attempts to fit the Eq. (1) of reference 1 to either offset or threshold values at different pressures when the measurements were first made obviously led to no sensible result, since thresholds are determined by clean-up and not by Eq. (1).

Once the Trichel pulse corona starts, the frequency of the pulses increases with increasing potential. Attempts to increase the frequency of the pulses by increase in potential to where the pulses merged were unsuccessful. Apparently, a spot was overloaded before merging occurred and a second spot appeared. With two spots active simultaneously, some interference was observed and the oscilloscope pattern became confused. Observation indicates that two spots acting at the same time can give Trichel pulses of different amplitudes. It is also not impossible that the offset potentials of two spots may not be the same. Thus, the oscilloscope will show additional fluctuations in current on a larger time scale with multiple spots acting. At low pressures with multiple spots the spots were spread rather randomly along the surface, but approaching well-spaced distribution owing to their mutual repulsion. At higher pressures spots appeared preferentially along one side of the wire. This could have been caused by slight lack of axial symmetry in the electrode system.

In air the general glow, as well as the spots, were usually visible at the threshold. At atmospheric pressure the glow was visible at 10 μa, and at 617 and 345 mm it was visible at only a few μa. In general, at lower pressures the luminosity appeared at somewhat lower currents. Overshoot at threshold was observed down to 100 mm. At the same pressure the current required for visibility was 0.2 μa. Spots associated with the Trichel pulses superposed on the general glow appeared as the potential was increased. There was always a dark space between spot and adjoining glow because of repulsion.[8] After the Trichel pulses appeared along with the glow, subsequent onsets had many bright spots superposed on the glow. It is possible that the discharge alternated between the two phases, glow and spots, and the composite picture seen was thus misleading. In some cases, when the spots migrated up and down the wire, the glow

[6] F. W. Peek, Jr., Trans. Am. Inst. Elec. Engrs. **31**, 1051 (1912).
[7] Loeb, Kip, Hudson, and Bennett, Phys. Rev. **60**, 714 (1941).

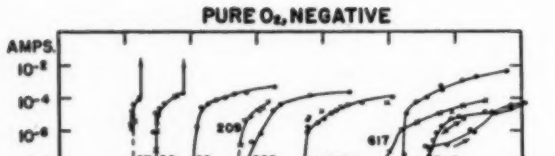


FIG. 9. Current-potential curves for negative wire corona in pure O_2. Note the very large pre-threshold currents and relatively small potential changes at transition at higher pressures.

followed always separated from the spot by a dark space, indicating simultaneous glow and spot discharges.

On occasions where the glow had not contracted to a spot and extinguished on decreasing potential, the dying glow persisted in diminishing intensity at spots showing no pulses. The conditioning of the surface by the apparently uniform glow must have left the surface unevenly conditioned as has been suspected from the formation of spots.

On occasion after standing the newly initiated discharge appeared as a uniform glow. Upon slightly increasing potential, random Trichel pulses appeared, presumably emanating from all portions of the wire but giving the visual appearance of a uniform glow. Further increase in potential caused the glow to change to a few spots that gave repetitive pulses instead of random ones. The reduction in pressure increased diffusion of phenomena reduced the local densities of bombardment and increased the energy of ion bombardment, so that localized clean-up and concentration gave way to more diffuse manifestations.

4. NEGATIVE WIRE IN PURE O_2

In pure O_2, as might be expected, the conditioning of the wire, which introduced the past history of the wire, complicated the picture as seen in the current-potential curves of Fig. 9. The first curve at 745 mm began with a wire that had been in mixtures with N_2. After running a discharge in pure O_2 the curve could be followed down back to 10^{-10} amp and appeared repeatable. A notable feature of the curves in O_2 is the steepness and continuity of rise from low values of 10^{-11} amp to the thresholds. The currents below threshold are by their magnitude and repeatability shown to be caused by a Townsend breakdown of low order that may start several hundred volts below the pulse or other thresholds. With the large pre-threshold current rise, the only indication at threshold was a relatively minor jump in current of the order of twofold and the *appearance of pulses* on the oscilloscope. This should be compared to the hundred-or-more-fold increase in air. At lower pressures the transitions were more marked. The pulse appearance was not accompanied by light emission owing to the peculiarly low visibility of the oxygen spectrum. It is strange that spots did not appear with Trichel pulses. The indications were that there was at first a fairly heavy uniform conditioning discharge down the length of the wire as indicated by the uniform distribution of current to guard rings and central collector. This distribution held even with the small increase in current and the appearance of random Trichel pulses. It appears as if the high concentration of O_2 deactivating the wire in competition with clean-up enabled Trichel pulses to form at a spot; but, having inadequate density of current, these pulses would go off during some extinction phase and reappear at some point along the wire. When, however, at higher current an anchored visible Trichel pulse appeared with a spot, the uniform glow and even distribution of current along the wire disappeared and current concentrated at the spot. Sometimes the uniform glow which occurred before the spot might persist, while the spot was active and visible, but usually not. It is also possible in those cases that the superposed glow could have been on during off periods of the spot. With the transition from the random pulses to fixed regular pulse, the current increase was by a factor of two. Strangely enough, except at 27-mm pressure, overshoot in potential was not observed. This is possibly caused by the heavy pre-threshold currents and relatively small jumps. At the lower pressure the effectiveness of the predischarge current is reduced because of the low density, thus leading to inadequate clean-up and overshoot.

One of the most puzzling features of the discharge in O_2 is the magnitude of the pre-threshold currents in pure O_2 and the lack of overshoot. It would almost appear that pure O_2 was much more transparent to photons acting on the cathode wire than air, as shown by the positive wire study.|| It must be noted that thresholds are about 500 to 1000 volts higher in O_2 than in air. Such conditions may be caused by greater difficulty of clean-up than with air. The predischarge must proceed by photoionisation *at the cathode* and *in the gas* as indicated by positive corona studies? The high potential leads to higher currents before clean-up and breakdown.

The Trichel pulses in pure O_2 are different from those in air in that they last from 1 to 5 microseconds, while in air they last in the order of 0.4μ sec as indicated by the synchroscope. It must be here stated that what the Dumont oscilloscope or the *synchroscope records* as a Trichel pulse is not necessarily the *actual ionizing current* but what the plane or cylindrical electrode senses electrostatically as the discharge proceeds. This has recently been pointed out by English.[8] Thus, conclusions here drawn may be in error, especially as regards the

|| As noted in the preceding paper on positive wire, heavy predischarge currents appear associated with the Townsend pulse type of predischarge currents identified by Lauer. The oscilloscope available for this study would not have resolved the pulses.

[8] W. N. English, Phys. Rev. **77**, 850 (1950).

amplitude of pulses, say, in point-to-plane and coaxial cylindrical geometry. The recorded *duration* of the pulses may not be entirely wrong in *relative* magnitude, though in actual time scale they appear to be overestimated. A longer duration to the Trichel pulse may be expected in O_2 if we surmise that ion formation and choking off the pulse by space charge proceeds more rapidly in air with its nitric oxides, etc., than in pure O_2, where the rather slow formation of O_2^- occurs.

In air it was impossible to follow the increase in frequency of Trichel pulses with increasing current and potential to the point where they merged because of multiple spot formation. In O_2 at *low pressures* the pulses followed each other so rapidly that, as soon as one pulse was extinguished, the other pulses succeeded them in regular time sequence. The increase in pulse rate observed was one in which the pulses followed one another with no appreciable interval between. The current increase at this point was such as to blow a 10-ma fuse. The space charge clearance in the case of the attenuated O_2 at low pressure must have been so rapid that the choking effect abruptly ceased and the discharge broke over to an arc. At higher pressures the space charge persists; and when a spot is overloaded, a new one forms as in air. In general, visible glows in negative corona in O_2 required the same current of 10 μa as did positive coronas. Only in single anchored spots did the visual appearance set in at slightly lower currents.

On decreasing the pressure the detailed phenomena changed in some respects. Thus, for instance, at 209 mm Trichel pulses marked the threshold at 2×10^{-9} amp and continued to 0.15 μa. These pulses were random but uniform in amplitude, a continuous ripple background current also being present from the Townsend discharge. Above 0.15 μa the pulses became more frequent but became non-uniform in amplitude. The visual glow appeared along the whole wire at 40 μa and was still visible at 200 μa.

The presence of the randomly spaced uniform amplitude pulses becoming more regular in time as potential increases and then going over to pulses of irregular amplitude appears not to pose any problems. The regular equi-amplitude pulses may be ascribed to a single anchored spot of fixed work function. The irregular ones could be ascribed to pulses from many different spots of irregular work functions. No evidence that such was the case is at hand. Numerous other hypotheses can be made but no final conclusion can be drawn.

With increasingly low pressures the evolution of the pulses, current, and spots underwent a number of abrupt and confusing transitions, some without further increase in current. As analysis and explanation are impossible at this point, it is not profitable to present the data.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

The Plastic Deformation of Pure Single Crystals of Lead and Copper*

PETER W. NEURATH† AND JAMES S. KOEHLER‡
Department of Physics, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received November 13, 1950)

Single crystals of 99.999 percent pure lead and copper were grown by the Bridgman method. A tensile creep apparatus is described. Measurements on copper at 23°C and at −190°C show that for resolved shearing stresses below 0.6 kg per mm² the steady-state creep rate is negligible (i.e., less than 10^{-6} per minute). The resolved yield stress for copper is 219±21 g per mm² at both the temperatures used. In the case of lead, for resolved shearing stresses below 0.13 kg per mm² the steady-state creep is negligible at −190°C. The yield stress at that temperature is 96±8 g per mm². Measurements were made on the steady-state creep of lead at 25°C and 110°C. The results produced by annealing were investigated.

INTRODUCTION

VERY few data are at present available on the plastic properties of pure single crystals of the face-centered cubic metals over a wide range of temperatures. The most completely investigated metal in this class is aluminum. The research to be described is an attempt to add information on the mechanical behavior of pure copper and lead crystals at small strains over a range of temperatures. The data obtained indicates that the behavior of copper at low strains is not what one would have expected in the light of previous data on aluminum. The experimental procedure is given in sufficient detail so that the experiments could be repeated elsewhere.

THE CREEP APPARATUS

The creep apparatus was designed to operate in the strain range from 2×10^{-6} to 4×10^{-2}. The specimen could be deformed and the load and strain measure-

* This research constitutes a portion of a thesis submitted by P. W. Neurath in partial fulfillment of the requirements for the degree of Doctor of Science at Carnegie Institute of Technology. This work was supported in part by ONR and NACA contracts.
† Westinghouse Fellow; present address, General Electric Company, Pittsfield, Massachusetts.
‡ Present address, Physics Department, University of Illinois, Urbana, Illinois.

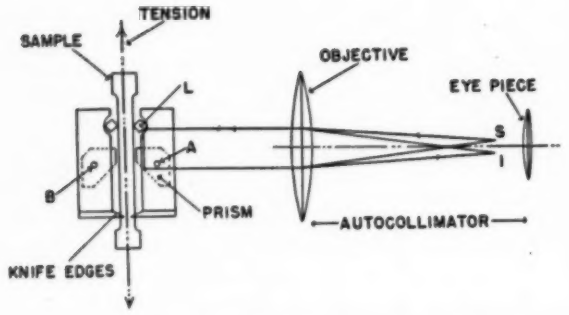


FIG. 1. The Tuckerman optical strain gauges and the viewing autocollimator showing the principle of operation.

ments could be obtained at specimen temperatures ranging from −190°C to 150°C. The load was increased by small but discrete amounts, and the strain was then observed as a function of time.

The schematic diagram (Fig. 1) shows the operation of the Tuckerman optical strain gauges and the associated autocollimator.§ S is an illuminated scale whose image I falls upon a second scale located in the focal plane of the objective. The position of the image depends on the angular position of the lozenge L. The image and the second scale constitute a vernier and are viewed through the eyepiece. A strain of 2×10^{-6} can be detected with this instrument. Two gauges were used to ensure that a strain indication results from an elongation and not from a straightening of an originally bent specimen. It is clear, however, that straightening which takes place in a plane perpendicular to the plane of the figure will give rise to an apparent extension, since both gauges will indicate the same positive strain. Difficulties associated with bending prevented the measurement of tensile strains smaller than 10^{-5}.

The gauges were mounted in two ways. First, they were pressed against the specimen by using two small coiled springs, one of which hooks over the pins at A and B of Fig. 1 and passes in front of the specimen. The other spring passes behind the specimen. Strain measurements made in this way checked well with the value predicted, using a known load on a polycrystalline specimen having a known Youngs modulus. It is difficult to mount an annealed single crystal in this fashion without cold-working it. The second mounting method used auxiliary supports for the gauges; the mechanical pressure between the gauges and specimen was made as small as possible. In general, this procedure produces data which is not reproducible at strains below 10^{-4}. The error occurs because the knife edges of the gauge slide along the specimen.

Figure 2 is a schematic diagram of the loading device, the cryostat, and the arrangement used to get light in and out of the cryostat. Loading was accomplished by sliding a weight along a lever, using a screw drive

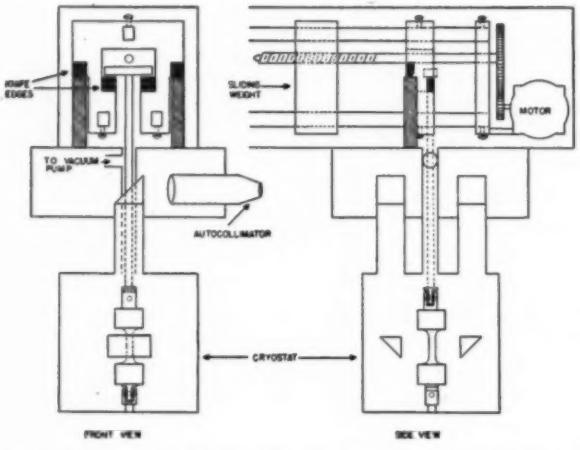


FIG. 2. General view of the creep apparatus, showing the loading lever, the cryostat, and the optical apparatus for strain measurement.

§ Produced by the American Instrument Company, Silver Springs, Maryland.

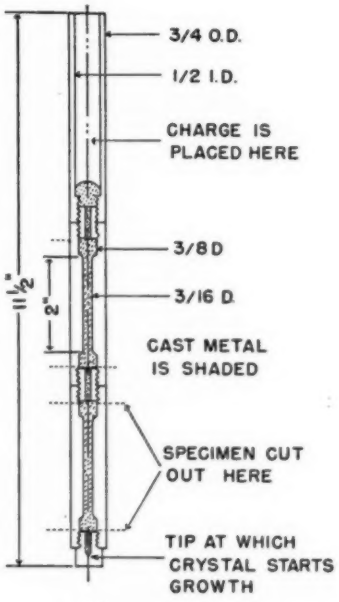


FIG. 3. Graphite crucibles used for the production of single crystals.

powered by a geared-down one rpm motor. Two interchangeable sliding weights were used, giving maximum loads of 40 kg (for copper) and 6 kg (for lead). Tungsten carbide knife edges and bearings are used. The sliding weights are counterbalanced in their zero position, and the load is measured by noting the position of the sliding weight. Universal joints above and below the specimen minimize non-axial loading. Split grips are used for easy specimen mounting. This loading system allowed load changes to be accomplished within a few minutes without the introduction of transient loads. There are two difficulties encountered with this loading system: First, friction does not allow significant measurements to be made at loads giving resolved shearing stresses less than 10 g per mm² on the 0.2-in. diameter specimens used by us. Second, the loading arrangement amplifies transient shocks present in the building. These

shocks produced transient stresses of the order of 10 g per mm² in the specimen.

The cryostat surrounded the sample and strain gauges. The light beams and thermocouple leads were brought out of the cryostat through closed tubes which extend above the surface of the coolant. Thermocouples were mounted in the grips just above and below the specimen. The cryostat and the loading mechanism were connected by a tube containing the loading rod, and the entire apparatus could be evacuated to avoid fogging of the optical system at low temperatures. Ice water and liquid nitrogen were used for cooling. Temperatures above 25°C were obtained by winding nichrome wire around the cryostat; the temperature was then maintained by a Micromax controller. A temperature gradient of one degree centigrade per inch of sample was tolerated.

PRODUCTION AND TREATMENT OF CRYSTALS

The single crystals were grown by the Bridgman method using *AGR* grade graphite crucibles of the type shown in Fig. 3. The purity of the original metals and of the resulting crystals is given in Table III. The

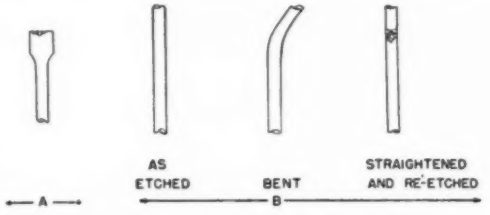


FIG. 4. The recrystallization of lead single crystals, A shows the usual situation in a tensile specimen, B illustrates rapid recrystallization after severe deformation.

charge was put into the top portion of the crucible after chemical cleaning. The following procedures were used to cast and grow the specimens:

(1) Lead samples *N*1 to *N*5 were cast by putting the graphite crucibles into a Pyrex tube which was connected to a vacuum system with a rubber hose. The samples were cast and outgassed at 450°C and 10^{-4} mm of mercury. The Pyrex tube was then sealed off and was finally lowered through the growing furnace (rate of lowering ½ in. per hour, 400°C).

(2) Lead samples *N*6 to *N*13 were all cast and grown simultaneously in a large stainless steel vacuum chamber. A traveling furnace surrounds the steel vacuum cylinder (rate 1 in. per hour, vacuum 5×10^{-4} mm, 425°C).

(3) Copper crystals were cast and grown in the vacuum furnace described by Siegel[1] (rate 4″ per hour, vacuum 5×10^{-5} mm, 1100°C).

The crystals were cut to length while still in the graphite. Three longitudinal cuts were then milled into the graphite crucibles, 120° apart, and the crucible could then be split to remove the specimen. The samples were then etched to make sure that they were single crystals.

The specimens were oriented to within ±1°, using Laue back reflection x-ray photographs. Some asterism and multiple spots were observed, particularly in the copper samples.

Since the graphite crucible does not contract as rapidly as the metal on cooling, it is to be expected that the samples will be deformed on cooling from their

[1] S. Siegel, Phys. Rev. **57**, 537 (1940).

TABLE I. Results for copper and lead crystals with negligible steady-state creep.

No. of crystal	Anneal before mounting	Temperature of run °C	Previous maximum stress g/mm²	Yield point g/mm²	Plastic modulus $\times10^{-6}$ dyne/cm²	
N 2a	none	0	?	535	10	
N 2a	none	24.6	770	880	12.6	
N 2b	600 v	0	?	265	7.2	
N 2b	none	0	425	455	10.8	
N 2b	none	24.6	702	700	12.7	
N 3a	none	0	?	460	8.0	
N 3a	600 v	0	724	600	11.2	
N 3b	600 v	0	?	425	9.6	
N 4a	600 v	0	?	460	5.3	
N 4a	none	0	670?	710	6.8	
N 4b	600 v	−190	?	367	11.3	
N 4b	600 v	0	587	435	12.2	
N 5a	600 v	0	?	325	12.0	
N 5a	600 v?	0	625	650	9.5	
N 5b	600 v	−190	?	335	11.8	
N 7a	1000 v	−190	?	241	5.6	
N 7a	600 v	23	372	297	8.9	copper
N 7b	1000 v	0	?	217	7.8	
N 7b	600 v	23	410	443	7.9	
N 7b	none	23	620	640	8.8	
N 8a	600 v	−190	?	290	6.2	
N 8a	none	0	362	380	6.5	
N 8b	600 v	−190	?	290	6.2	
N 8b	1000 v	0	455	225	9.5	
N 9a	1000 v	23	?	246	14.3	
N 9b	1000 v	0	?	225	15.6	
N 9b	none	23	521	540	16.3	
N 11a	1000 v	−190	?	184	5.3	
N 11b	1000 v	0	?	185	13.3	
N 11b	1000 H	0	400	200	7.0	
N 12a	1000 v, H	0	?	185	7.7	
N 12b	1000 v	−190	?	238	6.7	
N 3a		−190	?	118	1.7	
N 3b		−190	?	93	1.75	lead
N 4b		−190	?	80	2.7	
N 5b		−190	?	94.5	0.7	

All anneals lasted 20 hours. The second column shows their temperature in °C and whether done in a vacuum (v) or in a stream of pure, dry hydrogen (H). The stresses in the last three columns have all be calculated in the plane and direction of most probable slip.

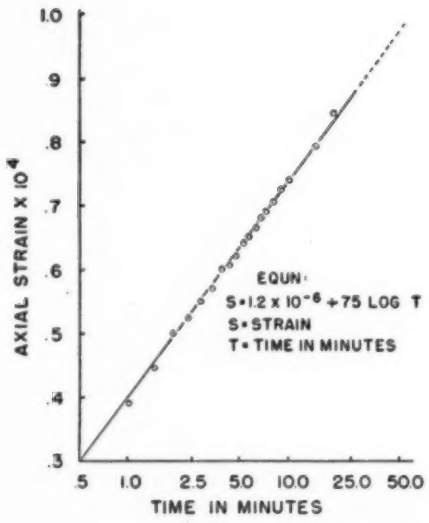


FIG. 5. Transient creep in a pure copper single crystal at 23°C under a resolved shearing stress of about 800 g/mm².

melting temperature. This strain amounts to about 1 percent for both lead and copper. In most cases, prolonged vacuum anneals were given before the specimen was tested. In some cases (see Table I) the samples were annealed in dry hydrogen.

No recrystallization of the specimens was observed during and after testing, although some recrystallization does occur at the cut ends of the lead crystals (see A of Fig. 4). A large inhomogeneous strain produces very rapid recrystallization in a pure lead crystal at room temperature. The events diagramed in Fig. 4B took place within a few minutes.

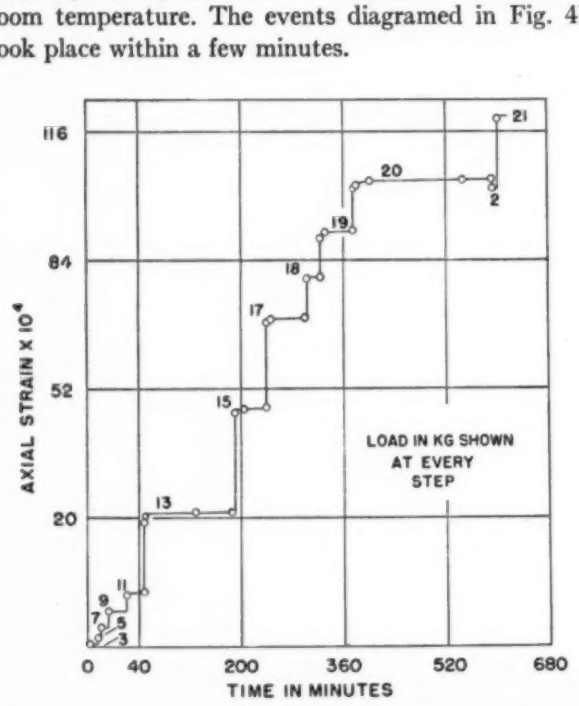


FIG. 6. Typical strain *versus* time curve for a hard specimen. Data obtained on a pure copper crystal at 0°C.

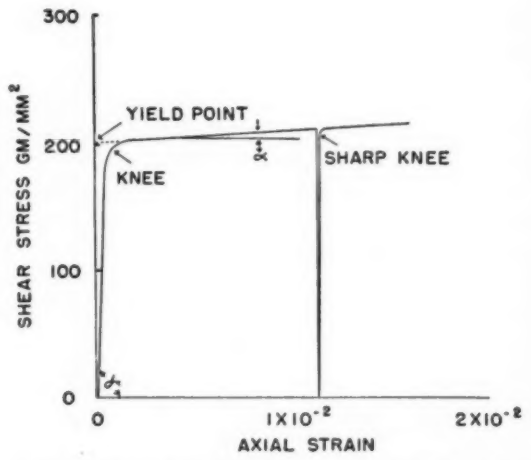


FIG. 7. Typical stress *versus* final strain curve for a hard specimen. Tanγ is the Young's modulus, tanα is the plastic modulus.

DATA

The mechanical behavior of the specimens can be separated into two classes depending on whether steady-state creep is or is not important. For copper crystals at −190°C and at 25°C and for lead at −190°C steady-state creep is negligible. The behavior of such "hard" specimens can be described as follows:

For a hard specimen, each increase in load is accompanied by a corresponding increase in length, which takes place within ten or twenty minutes. Figure 5 shows that the transient creep which occurs in a hard crystal increases logarithmically with the time. Figure 6 gives typical data on a copper crystal and shows clearly

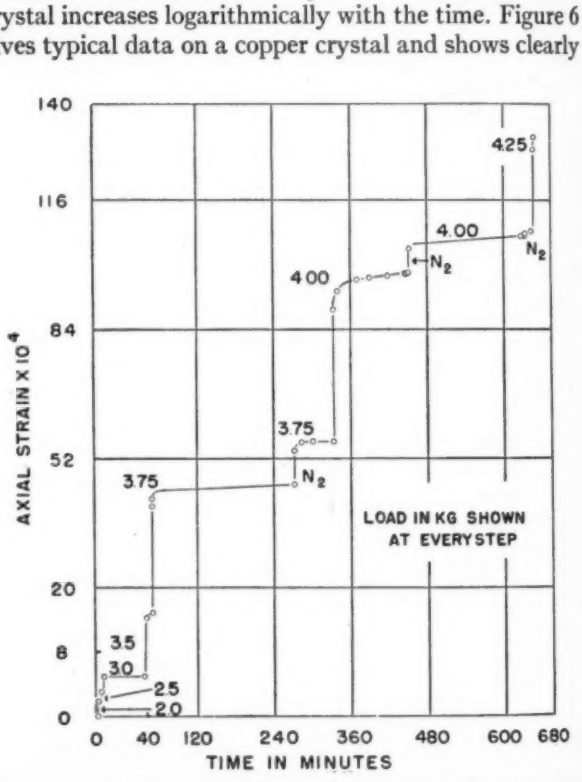


FIG. 8. Strain time curve for a pure lead crystal at −190°C. Liquid nitrogen added at points N_2.

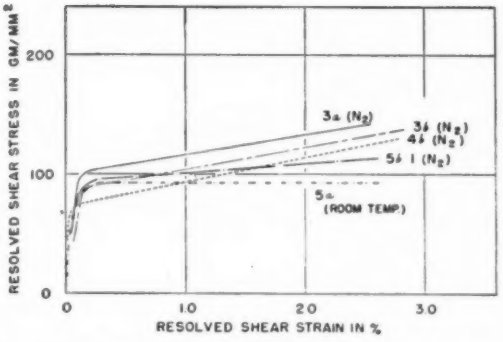


FIG. 9. Stress *versus* final strain curves for pure lead crystals. Curves N_2 run at $-190°C$. Specimen 5a run at room temperature.

that if the measurements are made slowly one can regard the strain as a function of the stress alone. In Fig. 7, we have, therefore, plotted a typical stress-strain curve of a hard specimen (the drawing is appropriate for a copper crystal). Note that for each stress the "final" strain reached after thirty minutes is plotted. The quantities of interest for a hard crystal are Youngs modulus (given by γ in Fig. 7), the plastic modulus (given by α), and the yield stress (given by the intercept of the plastic straight line with the stress axis as shown in Fig. 7). Typical strain time data for a lead crystal at $-190°C$ are shown in Fig. 8. The curve shows that the amount of steady-state creep is small although it is observable at large loads. The general appearance of the curve is the same as that found for copper, as illustrated in Fig. 6. Stress-strain curves for several lead crystals at $-190°C$ are shown in Fig. 9.

There are certain qualitative features of the behavior of hard specimens which should be mentioned here. First, the knee of the stress-strain curve (see Fig. 10) is rounded in a well-annealed specimen; in a crystal which has been deformed, the transition from elastic to plastic behavior is more abrupt (see Fig. 10, where the behavior on release and reloading is given). Second, there is actually a small amount of hysteresis on release and reloading which is not shown in Fig. 10. Figure 11 gives data on hysteresis in a copper crystal at room temperature. Read and Tyndall[2] have obtained similar data on zinc crystals.

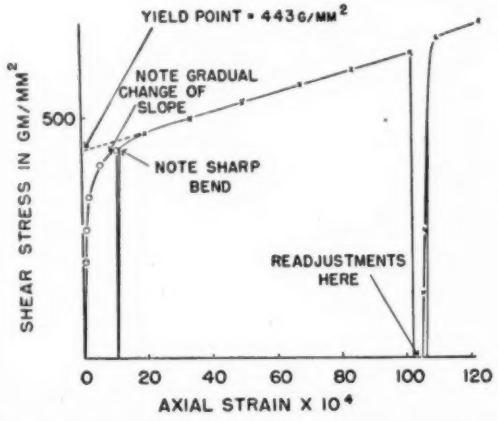


FIG. 10. Stress *versus* final strain for a pure copper crystal at 23°C after a 600°C vacuum anneal. The yield stress was about equal to the maximum prior stress.

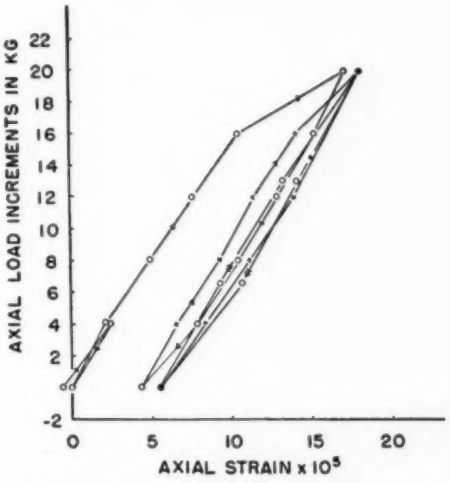


FIG. 11. Anelastic behavior in a pure copper crystal at 23°C.

A collection of the data obtained on copper crystals and on those lead crystals which were run at $-190°C$ is given in Table I. The following facts should be noted.

(1) Annealing at 600°C for 20 hours does not produce the same yield stress for all samples. The yield stress ranged from 280 to 460 g per mm² after this treatment.

(2) Annealing at 1000°C for 20 hours does appear to result in the same yield stress, i.e., 215 ± 21 g per mm².

(3) The yield stress appears to be the same at

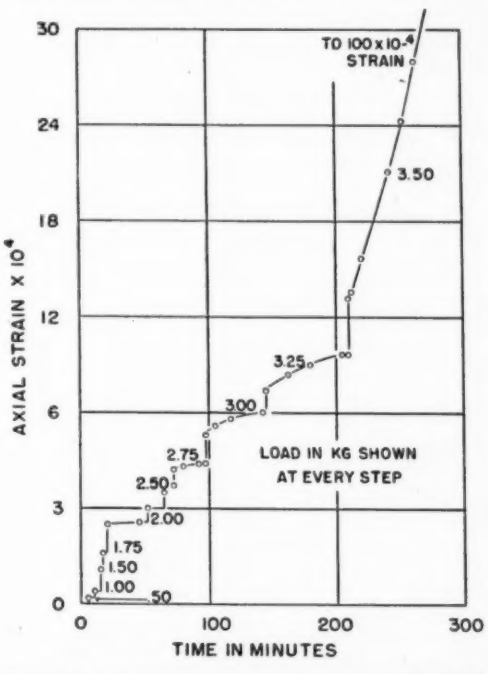


FIG. 12. Creep behavior of a pure lead crystal at 28°C.

[2] T. A. Read and E. P. T. Tyndall, J. Appl. Phys. **17**, 713 (1946).

TABLE II. Results for lead crystals showing steady-state creep.

No. of crystal	Anneal before mounting °C/time	Anneal after mounting	Temp. of run	Max. prior stress g/mm²	Minimum stress for a flow rate of $n \cdot 10^{-5}$/min for $n=1$	$n=10$	$n=100$
					grams per mm²		
N 5a			28	?	90?	96?	
N 6b			23.5	?	73	85	93
N 6b		24/0.3	24.2	93		105	112
N 6b		24/20	24.2	112	99	110	
N 6b	95/20g		24.7	111	69	85	
N 7a	130/24g	106/16	106	?	27		
N 7a		106/20	25	30	33	66	
N 7b	130/24g	60/20	26	?	51		
N 7b		62/15	62	60	42		
N 8a	220/24He		25	?	50	70	
N 8b		24/20	24.6	?	47	60	
N 8b	92/20g	none	23.9	67	58	67	
N 8b		23/20	22.3	67	67		
N 9a	130/24g	115/15	117.5	?	14	30	
N 9b	130/24g	118/18	118	?	17	30	33
N 10a	130/24g	120/12	120	?	20	30	35
N 10a		cooling	25.5	36	43		
N 10b	130/24g	120/15	25	?	25	51	
N 11a		100/14	100	?	17	31	
N 11a		100/24	25	31	34		
N 11b		120/12	25	?	45	60	
N 12x		120/60	123	?	14	20	
N 12x		cooling/60	24	22.4	22	34	

The second and third column list the annealing temperature in °C, the annealing time in hours, and whether done in glycol (g), helium (He), or air (no symbol), all in that order. The stresses in the last five columns have all been calculated in the plane and direction of the most probable slip.

−190°C as it is at room temperature. Comparing values after 1000°C V anneals, one finds, at −190°C, the yield stress is 221±25 g per mm², while at 0°C or 23°C it is 218±17 g per mm.

(4) Annealing in dry hydrogen at 1000°C produces, if anything, a lower yield point at 0°C than a vacuum anneal at 1000°C (i.e., 192 g per mm² after the hydrogen anneals).

(5) In general, prior tensile plastic strain raises both the yield stress and the plastic modulus.

(6) The liquid air temperature data for lead crystals (which had only room temperature heat treatment and, therefore, were not completely annealed) indicate a yield stress of 96±8 g per mm². Note that the plastic moduli for these lead crystals are appreciably lower than any of the values found for copper.

TABLE III. Specimen purity.*

a. Copper

Special high purity copper was supplied by the American Smelting and Refining Company. It was better than 99.999 percent pure (see J. S. Smart, Jr. and A. A. Smith, Jr., Trans. Am. Inst. Mining Met. Engrs. **143**, 272 (1941).

The analysis after growth in *AGR* graphite was Fe 0.00008 percent, Sn 0.00015 percent; Sb, Pb, Ni, Bi, Ag, Cd were looked for but not found.

b. Lead

Pig supplied by the National Lead Company; Bi 0.001 percent, Cu 0.0002 percent, Sb 0.001 percent, As 0.01 percent, Fe 0.0003 percent, Zn 0.02 percent, Ag 0.0015 percent; Sn, Ni, Cd were not found present.

Single crystal grown from American Smelting and Refining Company lead. Bi 0.0001 percent, Cu 0.0001 percent, Fe 0.0003 percent, Ag 0.0001 percent. No other impurities were found.

* We would like to thank Dr. McLellan of the American Smelting and Refining Company for the analyses quoted here.

In the case of lead measured at room temperature or above, steady-state creep was of importance. A typical creep curve obtained on a lead crystal at 28°C is shown in Fig. 12. From these data it is clear that at a given temperature the steady-state creep rate increases rapidly with increasing applied shearing stress. We therefore define the *creep limit* as the resolved shearing stress required to maintain a resolved shearing strain rate of 10^{-5} per minute (this rate is about ten times the smallest rate measurable with our apparatus). Data on the stress required to maintain strain rates of 10^{-4} and 10^{-3} per minute were also taken. One qualitative observation noted was that a transient increase in stress produced an increase in the strain rate during the presence of the large stress, but after the removal of the transient stress the rate was lower than the initial steady rate. Table II gives the steady-state creep data obtained using lead crystals. The following points should be noted.

(1) Room temperature anneals did not completely soften the material; the resulting creep limit was about 70 g per mm², whereas annealing at temperatures from 90°C to 120°C after the specimen has been mounted produces a creep limit of 34±5 g per mm² for a 25°C measurement.

(2) The creep limit of a well-annealed crystal decreases rapidly with increasing temperature. The creep limit at 110°C was 18±4 g per mm².

(3) The data indicate that the steady-state creep rate at a given temperature increases exponentially with stress.

Relation between Gas Evolution and Physical Properties of Liquids

V. G. SZEBEHELY*
Virginia Polytechnic Institute, Blacksburg, Virginia
(Received January 22, 1951)

Four heavy lubricating oils, three light lubricating oils, and three aircraft engine fuels were tested for viscosity, surface tension, and density, and a relation was established between the time rate of evolution and solution of air and the viscosity of the liquids. A definite relation was also found between solubility constants and half-lives as well as between half-lives of evolution and solution.

IN a recent paper[1] Dr. Schweitzer and the author reported on the investigation of the solubility and the rate of evolution of air in eighteen liquids. For further investigations, ten characteristic samples were selected. The measured values[2] of the physical properties of the samples are listed in Table I. One of the light lubricating oils (Number 7 in the table) is a synthetic oil, and one of the aircraft engine fuels (Number 8 in the table) is a special jet engine fuel. The previous investigations were extended also to distilled water which had entirely different solubility properties if compared to the hydrocarbons and therefore has not been included in the present paper.

Regarding solubility properties, three values are listed for each sample: the solubility constant (S), the half-life of evolution (T_E), and the half-life of solution (T_S). The half-life is defined as the time required to half complete the evolution or solution. The air-liquid volume ratio (r) is set at the value of four; i.e., the volume of the air section of the container is four times that of the volume of the liquid used. (See reference 1.)

The values of the viscosity, density, and surface tension are given at 70°F, since the solubility properties were determined at that average temperature. (It has been found that normal room temperature oscillations did not affect the solubility properties.)

In Fig. 1—using logarithmic scale for convenience of plotting—the relation is shown between half-life of solution (T_S) and solutility constant (S). The points form three groups. It can be seen that the points corresponding to the heavy lubricating oils (1, 2, 3, and 4) form the first group, the points of the light lubricating oils (5, 6, and 7) occupy the middle part of the curve, and the fuels tested form the third group. A similar curve is obtained if the half-life of evolution is chosen as dependent variable.

Figure 2 shows the relation between half-life of solution and half-life of evolution. The same grouping of

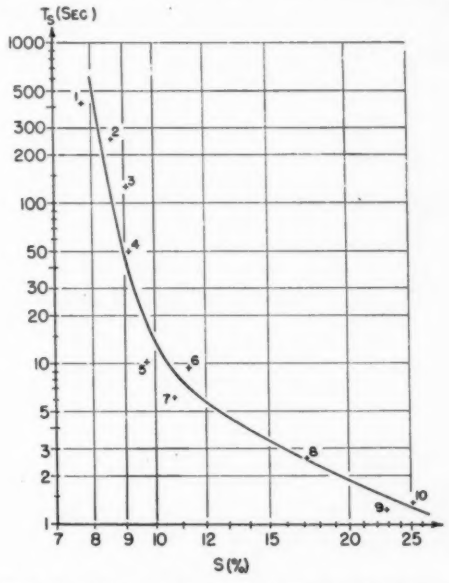


FIG. 1. Relation between half-life of solution and solubility constant.

TABLE I. Physical properties and experimental results.

	Kin. viscosity at 70°F (centistokes)	Density at 70°F (g/cm³)	Surface tension at 70°F (g/cm)	Solubility constant (%)	Half-life at $r=4$ for evolution (sec)	Half-life at $r=4$ for solution (sec)
1 Heavy lubricating oil	898	0.882	0.0299	7.75	51.4	414
2 Heavy lubricating oil	616	0.883	0.0308	8.62	34.8	252
3 Heavy lubricating oil	368	0.880	0.0301	9.05	31.0	125.4
4 Heavy lubricating oil	297.5	0.869	0.0293	9.11	17.9	50.2
5 Light lubricating oil	17.5	0.870	0.0274	9.70	4.74	10.22
6 Light lubricating oil	17.5	0.867	0.0286	11.30	7.63	9.46
7 Light lubricating oil	13.5	0.938	0.0259	10.72	3.56	6.14
8 Aircraft engine fuel	1.67	0.780	0.0239	17.20	0.236	2.595
9 Aircraft engine fuel	0.635	0.723	0.0191	22.80	0.128	1.214
10 Aircraft engine fuel	0.635	0.692	0.0181	25.14	0.137	1.344

* Associate Professor of Applied Mechanics, Virginia Polytechnic Institute.
[1] P. H. Schweitzer and V. G. Szebehely, J. Appl. Phys. 21, 1218–1224 (1950).
[2] The author is indebted to F. H. Fish, Jr., for measuring the densities and surface tensions of the samples.

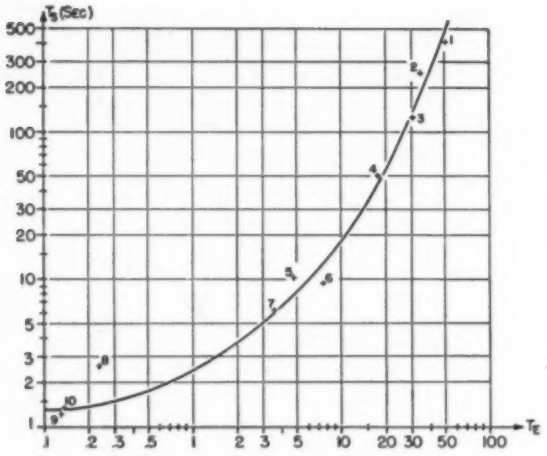


FIG. 2. Relation between half-life of solution and half-life of evolution.

the points can be seen with the exception of point 8. This point corresponds to the special jet fuel tested and shows a higher value of half-life for solution than expected, i.e., a relatively slow rate of solution.

It is interesting to note that the synthetic oil investigated (point 7 in the figures) shows no deviation from other liquids. Since surface tension, density, and viscosity did not vary independently, a relation was expected between one of these variables and the half-life of evolution. Selecting viscosity as representative variable, the following formula was obtained:

$$T_E = 0.263\nu^{0.81},$$

where ν is expressed in centistokes and T_E in seconds. The above formula was obtained by assuming a linear relation between $\log T_E$ and $\log\nu$ and the constants were determined by the method of least squares giving 0.975 for the correlation coefficient.

By the use of Fig. 2 one can connect viscosity with half-life of solution, and using Fig. 1 viscosity and solubility constant are connected.

It might be mentioned here that the rate of solution and evolution values reported in Table I were reproducible with an average error of 1–2 percent for the liquids numbered 1 to 7, and the error never did exceed 5 percent for the gasolines listed (8, 9, 10).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

The Electric and Magnetic Constants of Metallic Delay Media Containing Obstacles of Arbitrary Shape and Thickness*

SEYMOUR B. COHN
Sperry Gyroscope Company, Great Neck, New York
(Received December 4, 1950)

The dielectric constant and permeability of a metallic-obstacle medium are obtained in this paper by various static methods. The equivalent shunt capacitive susceptance and series inductive reactance of the individual obstacles are also determined. In the first part of the paper, a correspondence is established between an infinitely thin conducting obstacle and an aperture in an infinitely thin conducting wall. With the aid of this correspondence, delay medium formulas are given for several obstacle shapes. Formulas are then derived for the effect on the magnetic field of moderate obstacle thickness. Finally, it is shown how the electric and magnetic constants may be evaluated by electrolytic tank measurements in the case of obstacles having completely arbitrary shape and thickness. This has been done for a typical delay medium, and the effect of obstacle thickness on the index of refraction is shown.

INTRODUCTION

TWO basic analytical approaches have been used in the past to obtain the microwave index of refraction of metallic delay-lens media. With one approach[1] the electric and magnetic polarizabilities of the individual obstacles are first computed, and then the dielectric constant ϵ and permeability μ are evaluated by means of the following relations, which utilize rationalized mks units:

$$\epsilon = \epsilon_0 + \alpha_e N, \quad \mu = \mu_0 + \alpha_m N, \tag{1}$$

where N is the number of obstacles per unit volume, α_e and α_m are the electric and magnetic polarizabilities of the medium, and ϵ_0 and μ_0 the dielectric constant and permeability of the uniform medium in which the obstacles are placed. With ϵ and μ determined, the index of refraction may then be computed from $\eta = (\epsilon\mu/\epsilon_0\mu_0)^{\frac{1}{2}}$. This method gives good results if the obstacles are not too closely spaced and if the wavelength is considerably larger than the center-to-center spacings. The electric and magnetic polarizabilities have been obtained for several obstacle shapes by Kock,[1] Lengyel,[2] and Estrin.[3] The second approach involves the trans-

* The work described in this paper was accomplished for the Evans Signal Laboratory, Fort Monmouth, New Jersey, under Contract No. W36-039-sc-38250.

[1] W. E. Kock, Bell System Tech. J. **27**, 58 (1948).

[2] B. A. Lengyel, private communication, November 25, 1947.

[3] G. Estrin, J. Appl. Phys. **21**, 667 (1950).

mission-line equivalent circuit of the delay medium.[4–6] It is accurate for all frequencies and spacings if the constants of the equivalent circuit are correctly evaluated.

This paper will be principally concerned with static methods of obtaining the dielectric constant and permeability of the delay structure. It will first be shown that the electric and magnetic polarizabilities of an infinitely thin metallic obstacle are related, respectively, to the magnetic and electric polarizabilities of an identically shaped aperture in an infinitely thin conducting wall. Once this relationship is established, the dielectric constant and permeability of elliptical and other shapes of obstacles may be written down immediately. Next, a formula for the magnetic polarizability of a flat obstacle having moderate thickness will be derived for the case of the magnetic field parallel to the faces of the obstacle, and the equivalent circuit of the medium will be given. It will then be demonstrated how these constants may be determined for obstacles of any shape and thickness by an electrolytic tank method. Finally, the effect of obstacle thickness on the index of refraction of a typical delay medium will be shown.

By means of the methods just outlined the dielectric constant and permeability of a uniform array of obstacles may be determined for the three principal axes of the medium. With these parameters known, the behavior of a wave in the medium may be calculated by the methods previously utilized for optical waves in bi-refringent crystalline media.

RELATIONSHIP BETWEEN THE DIELECTRIC CONSTANT OF A DELAY MEDIUM AND THE MAGNETIC POLARIZABILITY OF AN APERTURE

In another article, this writer has derived by means of an application of Babinet's principle of electromagnetic duality the following formula for the average relative dielectric constant of a rectangular cell containing a single conducting obstacle at its center (Fig. 1):[7]

$$\epsilon_r = \epsilon/\epsilon_0 = 1 + (4M/abl), \tag{2}$$

where M is the magnetic polarizability of an aperture of the same shape and size as the obstacle in an infinitely thin conducting wall, and a, b, and l are the dimensions of the cell. As defined by Bethe,[8] M has the units length-cubed. The horizontal boundaries of the cell are electric walls, while the vertical boundaries are magnetic walls. The obstacle is assumed to be oriented so that one of its principal axes is parallel to the vertical incident electric field in the cell, while for the evaluation of M in the analogous case of the aperture the same corresponding principal axis is assumed parallel to the incident magnetic field.

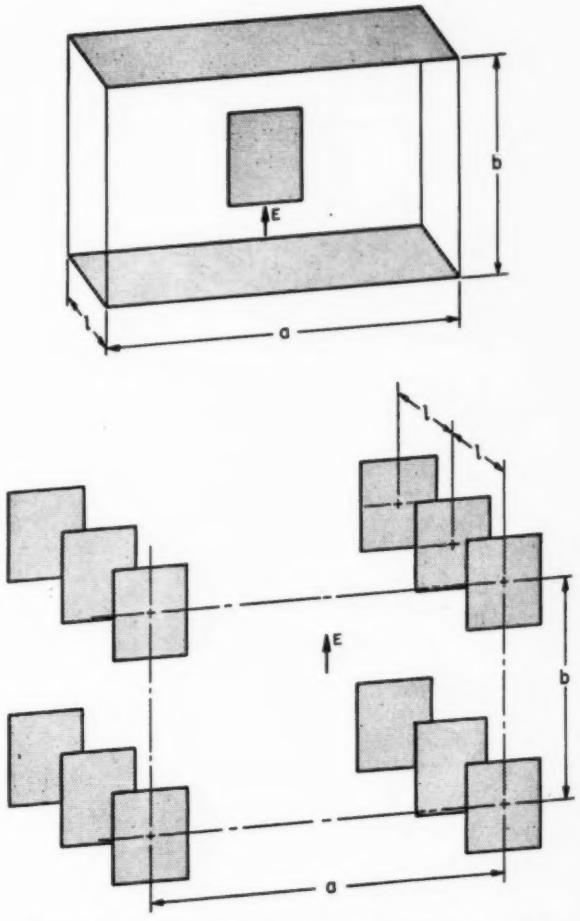


FIG. 1. The basic cell of a metallic delay medium and the corresponding array of obstacles.

It is apparent from symmetry considerations that the cell of Fig. 1 is a basic repeated unit of the array of obstacles shown in the same figure, and consequently the relative dielectric constant of the array is given by Eq. (2). Since the number of obstacles per unit volume is $N = 1/abl$, the relation can be rewritten more conveniently as follows:

$$\epsilon_r = 1 + 4MN. \tag{3}$$

A comparison of Eqs. (1) and (3) shows that the electric polarizability of a single obstacle is $\alpha_e = 4M\epsilon_0$. Although Eq. (3) was obtained specifically for an array of obstacles located at the intersections of a rectangular coordinate system, it applies equally well to any other kind of distribution.

With the aid of Eq. (3), the dielectric constant may be written down directly for a number of obstacle shapes previously unpublished. When the formulas

[4] S. B. Cohn, J. Appl. Phys. **20**, 257–262, 1011 (1949).
[5] S. B. Cohn, J. Appl. Phys. **21**, 674 (1950).
[6] J. Brown, J. Inst. Elec. Engr. **97**, Part III, 45 (1950).
[7] S. B. Cohn, "Determination of aperture parameters by electrolytic-tank measurements," Proc. Inst. Radio Engrs. (to be published.)
[8] H. A. Bethe, "Lumped constants for small irises," M.I.T. Radiation Laboratory Report 43-22, March 24, 1943. Also see *Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), Radiation Laboratory Series, Vol. 8, p. 178.

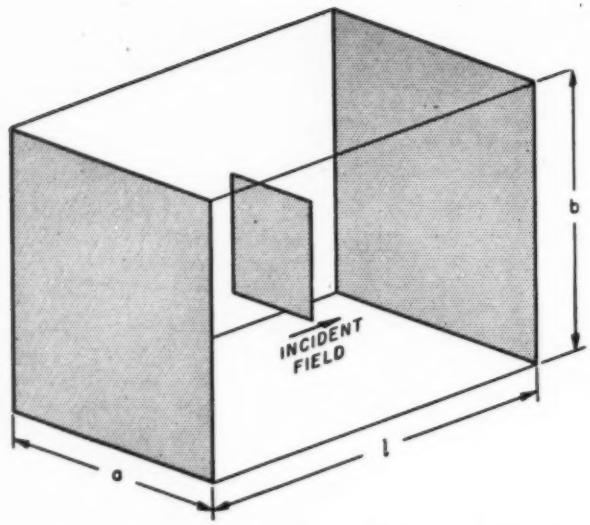


FIG. 2. Orientation of the field in a basic cell for the permeability derivation.

given by Bethe[8] for the magnetic polarizability of an elliptical aperture are used, one obtains the following dielectric constants for the incident field parallel, respectively, to the two axes of the infinitely thin elliptical obstacles:

$$\epsilon_r = 1 + \frac{4\pi ab^2k^2N}{3(1-k^2)(F(k)-E(k))} \quad \text{(ellipse, } E\,||\,a\text{)} \quad (4)$$

$$\epsilon_r = 1 + \frac{4\pi ab^2k^2N}{3[E(k)-(1-k^2)F(k)]} \quad \text{(ellipse, } E\,||\,b\text{)}, \quad (5)$$

where a is the semimajor and b the semiminor axis of the ellipse, $k=(1-(b/a)^2)^{\frac{1}{2}}$ is the eccentricity, and $F(k)$ and $E(k)$ are complete elliptic integrals of the first and second kind, respectively. These formulas for elliptical obstacles were previously derived by another method by B. Lengyel, but not published.[2] For the case of equal axes, Eqs. (4) and (5) reduce to the formula given by Kock for circular obstacles:

$$\epsilon_r = 1 + (16r^3N/3) \quad \text{(circle)}, \quad (6)$$

where r is the circle radius. Similarly, when the magnetic-polarizability formula for a long slit of constant width is used, the dielectric constant given by Kock for a metal-strip lens is obtained. The magnetic polarizabilities of rectangular, dumbbell, and many other aperture shapes have been obtained experimentally by a precise electrolytic tank method.[7] For example, the magnetic polarizability of a square was found to be $M=0.258d^3$, where d is the length of a side. The dielectric constant of a square-obstacle medium is, therefore,

$$\epsilon_r = 1 + 1.032d^3N \quad \text{(square)}. \quad (7)$$

RELATIONSHIP BETWEEN THE PERMEABILITY OF A DELAY MEDIUM AND THE ELECTRIC POLARIZABILITY OF AN APERTURE

The relative permeability of a delay medium consisting of an array of infinitely thin metallic obstacles that are all parallel to the same plane is equal to unity if the magnetic field is parallel to the obstacles, and is less than unity if the field is perpendicular to the obstacles. This statement is clearly true in the first case since the obstacles do not in any way affect the incident magnetic field, while in the second the field must follow a distorted path around the obstacles, and hence an average demagnetizing effect occurs. As in the case of the dielectric constant, the permeability is related to a basic parameter of an aperture, the parameter now being its electric polarizability. The relationship will be obtained with the aid of an analysis given in reference 7.

In the above mentioned reference it is demonstrated that the average dielectric constant of a rectangular cell containing a thin magnetic-wall obstacle at its center and perpendicular to the incident electric field (Fig. 2) is given by

$$\epsilon_r = 1 - (4P/abl), \quad (8)$$

where the electric polarizability P, as defined by Bethe,[8] has the dimensions length-cubed. Figure 2 shows a basic cell of this medium. The cell boundaries parallel to the electric field are magnetic walls, while those perpendicular are electric walls. Now by Babinet's principle, the boundary conditions will still be satisfied if all electric walls and magnetic walls are interchanged; the electric field replaced by a magnetic field, and the dielectric constant replaced by the permeability. These modifications thus change the cell into one containing a conducting obstacle in a perpendicular incident magnetic field. The relative permeability of the corresponding array of thin conducting obstacles is, therefore,

$$\mu_r = \mu/\mu_0 = 1 - 4PN. \quad (9)$$

The magnetic polarizability of a single obstacle is shown by Eq. (1) to be $\alpha_m = 4P\mu_0$. Equation (9) is independent of the way the obstacles are distributed as long as each obstacle is perpendicular to the incident uniform magnetic field.

When Bethe's formula for the electric polarizability of an elliptical aperture is inserted in Eq. (9), the following formula is obtained for the average permeability of an array of thin elliptical obstacles, all perpendicular to the incident magnetic field:

$$\mu_r = 1 - [4\pi ab^2N/3E(k)] \quad \text{(ellipse)}. \quad (10)$$

For the case of circular obstacles, this reduces to the following equation, which was first derived by Estrin:[3]

$$\mu_r = 1 - (8r^3N/3) \quad \text{(circle)}. \quad (11)$$

The electric polarizability of a long slit has also been derived by Bethe,[8] and hence the permeability of a medium containing thin narrow rectangles perpen-

dicular to the H field is as follows:

$$\mu_r = 1-(\pi w^2 lN/4) \quad \text{(rectangles)}, \tag{12}$$

where w is the width and l the length of the rectangles ($l \gg w$). For a metal-strip lens this becomes

$$\mu_r = 1-(\pi w^2 n/4) \quad \text{(strips)}, \tag{13}$$

where n is the number of strips per unit area normal to the strips.

Thus far, the electric polarizability of other shapes has not been determined. This parameter, and hence the permeability of a thin obstacle medium, may be measured with high precision, however, by an electrolytic tank method,[7] and it is hoped that such measurements will be made in the near future.†

PERMEABILITY OF MODERATELY THICK OBSTACLES

The preceding discussion applies only to infinitely thin obstacles perpendicular to the magnetic field. In the case of a medium containing thick flat obstacles parallel to the incident field, the relative permeability is less than unity by an amount that is often not negligible in practical delay-lens media. A formula for the permeability of thick obstacles that is valid for moderate thickness will now be derived utilizing an expression for the relative change in the resonant wavelength of a cavity due to a perturbation on its wall.[9,10] In mks units:

$$\frac{\Delta\lambda}{\lambda_0} \approx \frac{1}{2}\frac{\int_\tau (\epsilon|E|^2-\mu|H|^2)dV}{\int_V \mu|H|^2 dV}, \tag{14}$$

where E, H, and λ_0 are values for the unperturbed cavity, $\Delta\lambda$ is the wavelength change due to the perturbation, τ is the volume displaced by the perturbation, and V is the unperturbed cavity volume. In order for Eq. (14) to be valid, the projection of the perturbation into the cavity must be much smaller than its dimensions parallel to the cavity wall. The formula is exact in the limit of vanishing projection.

Assume a thick obstacle delay structure whose obstacles each have the volume τ and that are located at the intersections of a rectangular coordinate system. By symmetry considerations select a basic rectangular cell as shown in Fig. 3(a), and assume an incident uniform magnetic field in the direction indicated. Let the boundaries parallel to H be electric walls, and those perpendicular to H be magnetic walls. Also, let the cell be resonated by a large capacitance in series with an electric wall at a low enough frequency so that $E \approx 0$ in the cell (Fig. 3(b)). The two half-thickness obstacles projecting into the cavity are the perturbations whose effect is to be determined. Since the unperturbed H field in the cavity is uniform throughout the cavity, and $E \approx 0$, Eq. (14) may be evaluated as follows:

$$\Delta\lambda/\lambda_0 \approx -\tfrac{1}{2}\tau/V. \tag{15}$$

Let L be the unperturbed inductance of the cell, and ΔL the change in inductance due to the perturbation. The average relative permeability of the medium is related to L and ΔL by

$$\mu_r = (L+\Delta L)/L = 1+(\Delta L/L). \tag{16}$$

But $\lambda_0 \propto (LC)^{\frac{1}{2}}$, and hence

$$\Delta\lambda/\lambda_0 \approx \tfrac{1}{2}\Delta L/L. \tag{17}$$

Equations (15)–(17) give

$$\mu_r \approx 1-(\tau/V).$$

The number of obstacles per unit volume is $N = 1/V$ and therefore,

$$\mu_r = 1-\tau N. \tag{18}$$

Equation (18) is valid for any distribution of obstacles of any shape if the obstacles are of moderate thickness and are parallel to the incident H field. For a metal-strip medium Eq. 18 is more conveniently written in the following form:

$$\mu_r = 1-wln \tag{19}$$

where l is the thickness of the strip, and w and n are as previously defined.

The magnetic polarizability of a single moderately thick conducting obstacle is seen by comparison of Eqs. (1) and (18) to be $\alpha_m = -\tau\mu_0$. It may be shown in other ways that α_m is twice this value for a long conducting circular cylinder whose axis is perpendicular

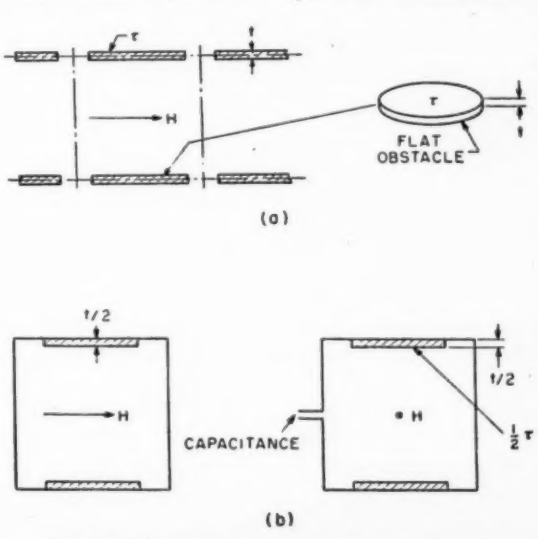


FIG. 3. Constructions for the thick obstacle analysis.

† *Note added in proof.*—The electric polarizability of a square aperture has been determined by electrolytic tank measurement to be $P = 0.1137d^3$, and hence the permeability of a square-obstacle medium is $\mu_r = 1-0.455d^3N$.

[9] J. C. Slater, "Forced oscillations in cavity resonators," M.I.T. Radiation Laboratory Report 43-16, December 31, 1942; Eq. VII; 22.

[10] W. W. Hansen and R. F. Post, J. Appl. Phys. **19**, 1059 (1948).

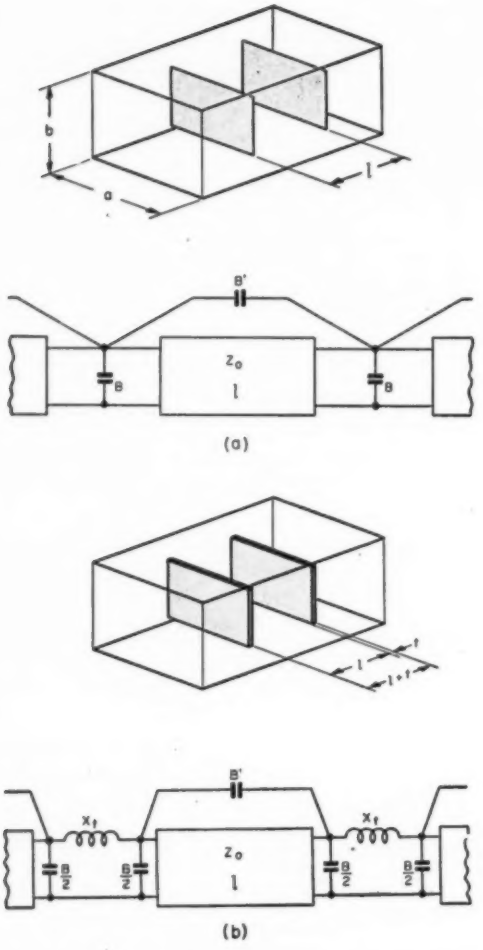


FIG. 4. (a) The equivalent transmission line and circuit for infinitely thin obstacles and (b) the equivalent transmission line and circuit for thick obstacles.

to the incident H field, and is three-halves times this value for a conducting sphere.

In order to compute the index of refraction for large wavelength, it is necessary to know the dielectric constant of the thick obstacle medium. A simple formula cannot be derived in this case, but it will be shown later in this paper how the dielectric constant of a thick obstacle medium may be measured accurately by an electrolytic tank method.

EQUIVALENT CIRCUIT OF A THICK OBSTACLE MEDIUM

The electric and magnetic effects of the obstacles may be considered as though they are averaged over the medium only if the wavelength is large enough so that $d\eta/df$ is negligible. When the spacing l is an appreciable part of a wavelength, however, this procedure no longer gives accurate results, and the equivalent-circuit method is more useful. Figure 4(a) shows the equivalent transmission line and the equivalent circuit for a medium containing infinitely thin obstacles,[4,5] while Fig. 4(b) shows the equivalent transmission line and circuit for thick obstacles. The boxes represent the Pi-section (or T-section) equivalent of a uniform line of length l and characteristic impedance Z_0. It can be seen that the obstacle thickness necessitates a new element, X_t, in the circuit. A formula for this element will now be derived for moderate thickness.

Let the inductance per unit length of the uniform equivalent parallel-plane line be L_0. Then, since the total length of one section of the loaded line is $l+t$, the inductance of one section with the obstacles removed is $L=(l+t)L_0$. The inductance with the obstacles inserted is $L+\Delta L$. Hence, the inductance of the length of line l between adjacent obstacles is $L_0 l$, and the inductance of the length of line t containing an obstacle is $L_t=L_0 t+\Delta L$. When Eqs. (16) and (18) and the relations $X_t=\omega L_t$, $N=1/ab(l+t)$, and $\omega L_0=2\pi Z_0/\lambda$ are substituted, one obtains

$$X_t/Z_0=(2\pi t/\lambda)(1-\tau/abt). \tag{20}$$

If the volume abt contains more than one obstacle, as would be the case with a hexagonal distribution of

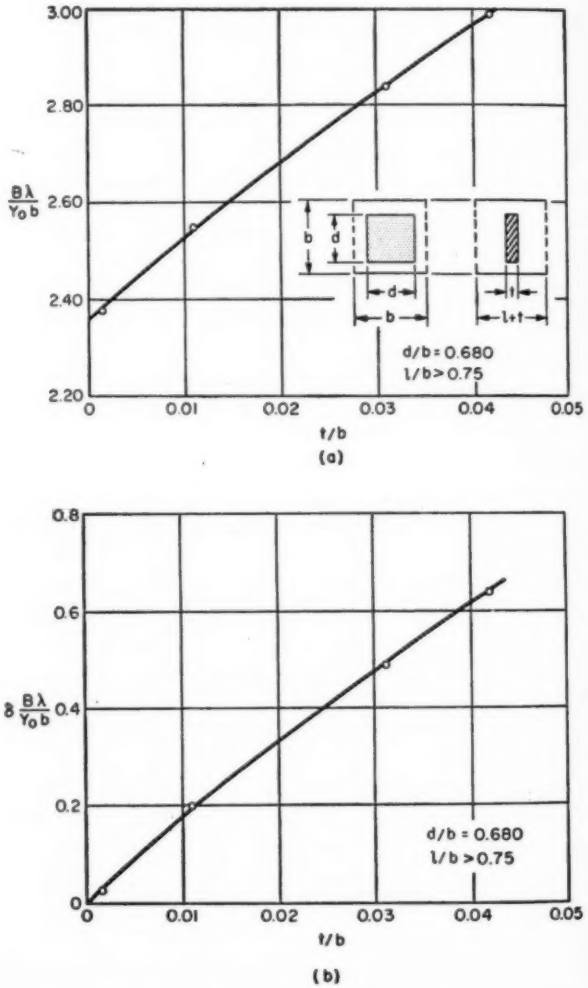


FIG. 5. The susceptance (a) and the increase of susceptance (b) of a square obstacle as a function of thickness.

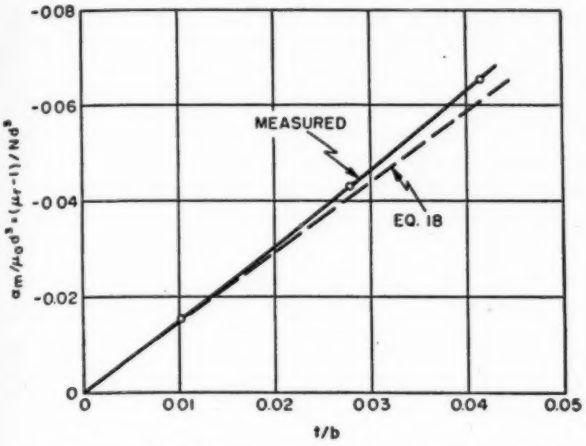


FIG. 6. Effect of thickness on the magnetic polarizability of a square obstacle for the incident magnetic field parallel to the obstacles.

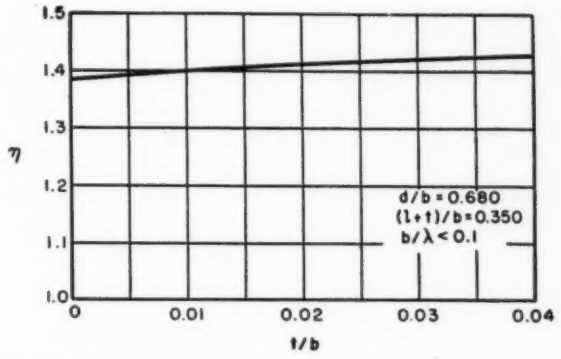


FIG. 7. Effect of thickness on the low frequency index of refraction of a medium containing square obstacles.

disks, for example, then Eq. (20) is still valid if τ is taken to be the total volume of the obstacles in abt.

The shunt capacitive susceptance B of a thick obstacle may be determined experimentally by the method to be described in the next section. The bridging susceptance B' should not be affected appreciably by the obstacle thickness, and hence values of B' for zero thickness may be used, if known. In the case of the metal-strip medium, accurate formulas are available[4] for the open- and short-circuit susceptance of a half section, and from these B' is readily determinable. It is a simpler procedure, however, to alter the open- and short-circuit susceptances to take account of X_t and the change in B, and then to compute the index of refraction and characteristic admittance of the medium directly from these modified parameters.

ELECTROLYTIC TANK MEASUREMENT OF THICK OBSTACLES

The dielectric constant and permeability of a medium containing obstacles of any shape and thickness may be measured accurately by an electrolytic tank method. The dielectric constant measurements are made on a conducting model of the obstacle, while the permeability measurements are made on a nonconducting model. The principles of the method will now be described. The actual measurement techniques and apparatus are those discussed in references 5 and 7.

First assume a medium that can be reduced by symmetry considerations to a single rectangular cell containing one or more obstacles. An electrolytic model of the cell may then be made in which current flow lines represent the electric field lines of free space. Then, because of the analogy between displacement flux in space and current in an electrolyte, the average relative dielectric constant of the delay medium is equal to the ratio of the conductances of the cell with the obstacle (or obstacles) inserted to that with the obstacle (or obstacles) removed. If the volume of the obstacles causes an appreciable change in the height of the solution, both conductances must be made to apply to the same water height as follows. Let G_1 and h_1 be the conductance and water height with the obstacles inserted, and G_2 and h_2 the corresponding quantities with the obstacles removed. Then G_2h_1/h_2 would be the conductance of the cell without the obstacles if the water height were h_1. Hence, ϵ_r is given by

$$\epsilon_r = G_1h_2/G_2h_1. \tag{21}$$

The shunt susceptance B of the obstacles may be computed from this formula by using $B=(\epsilon_r-1)B_2$, where B_2 is the capacitive susceptance of the cell without obstacles.

Similarly, the permeability of the medium may be obtained by measurements on a nonconducting model of the obstacle in the cell. In this case the current flowing by the nonconducting obstacle represents perfectly the magnetic field surrounding a conducting obstacle having zero skin depth. The nonconducting obstacle must, of course, be oriented with respect to the current the same way that its counterpart is oriented with respect to the magnetic field in space. If G_1, G_2, h_1, h_2 are defined as above, μ_r is given by

$$\mu_r = G_1h_2/G_2h_1. \tag{22}$$

Equations (21) and (22) take full account of proximity effects if the cell is the exact basic unit of the entire medium. If the obstacles are widely enough spaced so that proximity can be neglected, the following relation may be used to compute ϵ_r and μ_r for any distribution of the obstacles, once ϵ_r and μ_r are determined for one distribution:

$$\epsilon_r - 1 \propto N, \quad \mu_r - 1 \propto N. \tag{23}$$

MEASURED DATA FOR THICK OBSTACLES

Electrolytic tank tests have been made for the array of thick square obstacles whose basic cell is shown in Fig. 5(a). The measured obstacle susceptance *vs* thickness is plotted in Fig. 5(a) for the case of $l/b>0.75$, in which range B is independent of l/b. The change in B from the zero thickness value is plotted in Fig. 5(b).

This curve should be the same for any l/b as long as $t \ll l$. In Fig. 6, a curve giving the magnetic polarizability of a single obstacle and the permeability of an array of obstacles is plotted as a function of t. It is seen that excellent agreement with Eq. (18) is obtained for $t \ll b$.

Figure 7 shows the variation with obstacle thickness of the low frequency index of refraction of a particular square-obstacle medium. The following equation was used for the computation:

$$\eta = (\mu_r \epsilon_r)^{\frac{1}{2}} = \left[\mu_r \left(1 + \frac{B\lambda}{2\pi Y_0 (l+t)}\right)\right]^{\frac{1}{2}}. \quad (24)$$

The susceptance B increases with t while μ_r decreases, but since the change in B predominates, a net increase in η occurs. In computing μ_r, the measured curve of Fig. 6 was used. B was set equal to the increment plotted in Fig. 5 plus the susceptance of a zero thickness obstacle having the same ratios of d/b and l/b. The latter was obtained from Fig. 7 of reference 5.

The change in η is seen to be quite small in Fig. 7, if the thickness is moderate. For example, if $b = 1.000$ inch and $t = 0.010$ inch, the increase in η is only about one percent. If, however, only B were affected by thickness and not μ_r, then the increase would have been two percent.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Lattice and Grain Boundary Self-Diffusion in Silver*

R. E. HOFFMAN AND D. TURNBULL
General Electric Research Laboratory, Schenectady, New York
(Received November 22, 1950)

The lattice and grain boundary self-diffusion coefficients of high purity silver have been measured over extended temperature ranges. It has been found that the lattice diffusion coefficients are described by the equation,

$$D_L = 0.895 \exp(-49500/RT) \text{ cm}^2 \text{ sec}^{-1}$$

and are independent of the thermal history of the crystals within wide limits. The absolute grain boundary diffusion coefficients fit the equation,

$$D_B = 0.03 \exp(-20{,}200/RT) \text{ cm}^2 \text{ sec}^{-1}$$

and are independent of the grain size of the specimens.

INTRODUCTION

FREQUENTLY the coefficients of self-diffusion measured by different investigators fail to agree. For example, measured activation energies for self-diffusion in copper range from 45 to 61 kcal/g-atom.[1–4] There are many factors which could affect the measured value of the diffusion coefficient and cause these variations. It was the aim of this investigation to study quantitatively the effect on the self-diffusion coefficient of silver of two such factors, *viz.*, thermal history and grain boundary diffusion.

It is a widely accepted theory that self-diffusion in face-centered cubic metals takes place by a lattice vacancy mechanism.[5] Since the equilibrium concentration of vacancies decreases sharply with temperature, it is conceivable that a crystal could become supersaturated with vacancies after a rapid quench from a high temperature. According to the vacancy mechanism the lattice diffusion coefficient, D_L, is directly proportional to the concentration of vacancies. Hence, D_L should be larger for a specimen that is cooled rapidly to some temperature, T, than for one that is slowly cooled to the same temperature as long as the vacancy supersaturation persists.

The earliest measurements of the rate of grain boundary diffusion in metal systems appear to have been made on the diffusion of thorium in polycrystalline tungsten. Dushman and Koller[6] and Clausing[7] demonstrated that the rate of diffusion of thorium to the surface of a tungsten wire increased with decreasing grain size of the wire. Langmuir[8] has summarized and interpreted these results. He calculated that the activation energy for diffusion of thorium in tungsten is 90 kcal/g-atom at tungsten grain boundaries and 120 kcal/g-atom in the tungsten lattice.

More recently Achter and Smoluchowski[9] have found that the activation energy for the diffusion of silver along copper grain boundaries separating crystals hav-

* This work was supported by the U. S. AEC under Contract No. W-31-109-Eng-52.

[1] Steigman, Shockley, and Nix, Phys. Rev. **56**, 13 (1939).
[2] B. V. Rollin, Phys. Rev. **55**, 231 (1939).
[3] Raynor, Thomassen, and Rouse, Trans. Am. Soc. Metals, **30**, 313 (1942).
[4] M. S. Maier and H. R. Nelson, Trans. Am. Inst. Mining Met. Engrs. **147**, 39 (1942).
[5] C. E. Birchenall, Am. Soc. Metals Seminar on "Atom movements," Chicago, 1950.
[6] S. Dushman and L. R. Koller, unpublished research (1924).
[7] P. Clausing, Physica **7**, 193 (1927).
[8] I. Langmuir, J. Franklin Inst. **217**, 543 (1934).
[9] M. R. Achter and R. Smoluchowski, Phys. Rev. **76**, 470A (1949).

ing a common crystallographic axis [100] is of the order of one-half the activation energy for diffusion of silver through copper crystals. They have also found that the rate of grain-boundary penetration of silver into copper decreases as the difference in relative orientation of adjacent grains decreases.

Recently the rates of self-diffusion of radioactive silver into single-crystalline and fine-grained polycrystalline silver have been compared.[10] The diffusion coefficients, D_L, of single crystals were described satisfactorily over the entire temperature range of the investigation (500°C to 900°C) by the same relationship as that used by W. A. Johnson[11] to fit his data on the rate of self-diffusion in coarse-grained silver for the temperature range 725°C to 950°C. This relation is

$$D_L = 0.895 \exp(-45950/RT) \text{ cm}^2 \text{ sec}^{-1}. \quad (1)$$

At temperatures exceeding 700°C the apparent diffusion coefficients, D_B', for fine-grained polycrystals[10] also were described by Eq. (1), but in the range 500° to 700°C they were larger by a factor of 2 to 6 and fit the equation

$$D_B' = 2.3(10)^{-5} \exp(-26400/RT) \text{ cm}^2 \text{ sec}^{-1}. \quad (2)$$

Whether or not the measured diffusion coefficient, D_B', differs from D_L depends on the ratio of the absolute grain-boundary diffusion coefficient to the lattice diffusion coefficient, D_B/D_L. It has been shown elsewhere[12] that D_B' will not be significantly different from D_L unless $D_B/D_L > N$, where N is a number that is some function of the grain size. (N is always larger than unity and under certain conditions is of the order of 10^4 to 10^5 for a grain size of 0.05 cm.) Assuming that the activation energy for grain boundary diffusion is less than that for lattice diffusion, it is clear that D_B/D_L must increase with decreasing temperature. The above results can then be explained by postulating that for the experimental conditions used, $D_B/D_L < N$ at temperatures above 700°C, so that $D_B' = D_L$. At lower temperatures $D_B/D_L > N$, and grain boundary effects become evident.

Since D_B' values were calculated on the assumption of uniform penetration, it is evident that they are not the absolute grain boundary diffusion coefficients and it is not clear whether or not the activation energy, 26.4 kcal/g-atom, can be identified with that of grain boundary self-diffusion.

Fisher[13] has analyzed the problem of penetration into a polycrystalline specimen for conditions such that the transport of matter along grain boundaries is important. Making the assumptions (a) that the grain boundary region is a uniform zone of thickness δ with a constant concentration of isotope at a given distance from the plated surface and (b) that the diffusion outside the boundaries is primarily normal to them, he derived a method whereby the ratio D_L/D_B could be calculated by numerical analysis of data on the penetration of isotope. Application of the numerical analysis to diffusion data showed that to a good approximation the concentration within the grain boundary region remained essentially constant throughout the period of the diffusion anneal with respect to diffusion outside it and that also in the boundary region $\partial C/\partial t = 0$ at the time of measurement, where C is the isotope concentration and t the time. With these assumptions a simple analytical expression for $D_L/D_B\delta$ was obtained as follows:

$$D_L/D_B\delta = [d(\log a)/dy]^2(\pi D_L t)^{\frac{1}{2}}/2(\log e)^2, \quad (3)$$

where a is the activity of the isotope per unit volume at a distance y measured in a direction normal to the free surface on which the isotope is deposited prior to the diffusion anneal. If a value of δ is assumed, the absolute magnitude of D_B can be calculated from activity-penetration data with the use of Eq. (3).

TABLE I. Reduction schedule for polycrystals.

	Series 1	Series 2	Series 3
Casting	1.00″	1.00″	0.625″
First reduction	0.750″ (44%)	0.800″ (36%)	0.525″ (30%)
Recrystallization[a]	1 hr—600°C	2 hr—600°C	2 hr—600°C
Second reduction	0.500″ (55%)	0.500″ (60%)	0.450″ (26%)
Recrystallization	1 hr—600°C	2 hr—600°C	2 hr—600°C
Third reduction	0.375″ (43%)	0.340″ (54%)	0.372″ (32%)
Final recrystallization	2 hr—600°C	2 hr—600°C	1 hr—600°C[b]

[a] In Series 1 and 2, recrystallization took place in N_2, in Series 3, in vacuum.
[b] The final recrystallization of Series 3 was performed after the rod had been cut into ¼″ sections.

Earlier data[10] on the self-diffusion of silver into polycrystalline specimens at relatively low temperatures were obtained by a surface counting technique and so could not be used, without the aid of arbitrary assumptions, to calculate the absolute values of D_B. This paper will describe the earlier experiments in more detail and disclose the results of new experiments in which activity-penetration curves were obtained as a function of temperature for the self diffusion of silver into fine-grained polycrystalline specimens in the range 375° to 500°C. From these data absolute values of D_B are calculated with the aid of Fisher's analysis.

EXPERIMENTAL

A. Specimen Preparation

Two grades of silver were used. One was obtained from Handy and Harmon Company (H-H silver) and contained less than 0.03 percent by weight metallic impurities. The other, containing less than 0.001 percent metallic impurity, was obtained from the Johnson-Matthey Company (J-M silver).

[10] D. Turnbull, Phys. Rev. 76, 471A (1949).
[11] W. A. Johnson, Trans. Am. Inst. Mining Met. Engrs. 143, 107 (1941).
[12] D. Turnbull, Am. Soc. Metals Seminar on "Atom movements," Chicago, 1950.
[13] J. C. Fisher, J. Appl. Phys. 22, 74 (1951).

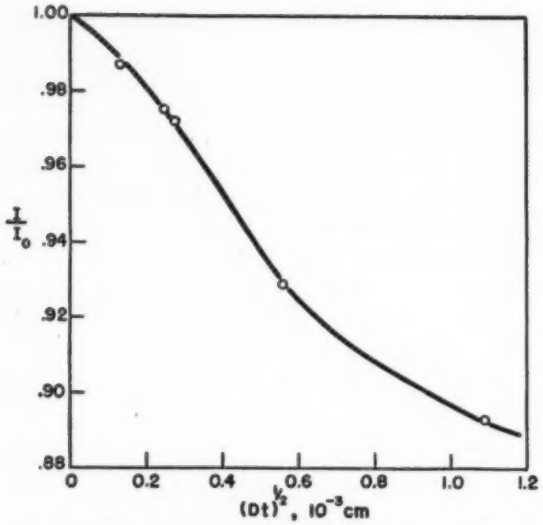


FIG. 1. Empirical curve for the decrease in activity as a function of $(Dt)^{\frac{1}{2}}$, measured at 724°C.

A single crystal of H-H silver, $\frac{1}{2}''$ in diameter, was grown in a graphite mold by the Bridgmen technique in a hydrogen atmosphere. Cylindrical specimens $\frac{1}{4}''$ long were carefully cut from the crystal with a jeweler's saw and the faces machined to be perpendicular to the specimen axis within an angle of ten minutes. After 5 mils were removed from their surface by etching, the specimens were heated in vacuum (10^{-5} to 10^{-6} mm) according to the following schedule: 1 hr at 250°C, 1 hr at 350°C, and 2 hr at 450°C. This treatment served to remove hydrogen and to cause recovery from residual strains.

The single-crystal specimens used to measure the effect of thermal history were first heated for one hour at 900°C in evacuated quartz tubes. One set of these was removed from the furnace and air-cooled to room temperature, another set was furnace-cooled to room temperature at the controlled rate of 30° per hour.

For all except one series of experiments on polycrystalline material, the stock H-H silver was melted, cast, and then alternately swaged and recrystallized until the desired grain size was obtained. The detailed steps of the process are shown in Table I. In the series of experiments on the high purity J-M silver (series 4), the specimens were cut directly from the $\frac{1}{4}''$ rod as received from the supplier and annealed at 700°C for 15 minutes.

B. Electrodeposition of Radioactive Silver

Radioactive Ag^{110} was deposited on one of the circular faces of each specimen by electrolysis from a cyanide bath. The silver in the electrolyte was composed of one part of radiosilver and one hundred parts inactive silver. After electrolysis the specimen was thoroughly washed and the coating rubbed with a cloth in order to remove any loosely adherent parts.

When diffusion experiments are carried out under conditions where evaporation of the specimen may be appreciable, it has been considered necessary to "sandwich" the isotope between two layers of the inactive material. W. A. Johnson[11] formed the sandwich by pressure welding two disks, each plated with isotope. In the experiments to be described the sandwich technique was used for some specimens; but when it was found that the diffusion coefficients measured for the sandwiched specimens did not differ perceptibly from those measured on single disks with the isotope at a free surface, most of the experiments were carried out on the latter type of specimen.

After electroplating, the specimens were sealed in evacuated Pyrex or quartz tubes.

C. Diffusion Anneal

The diffusion anneal was carried out in a nichrome wound furnace having a uniform controlled tempera-

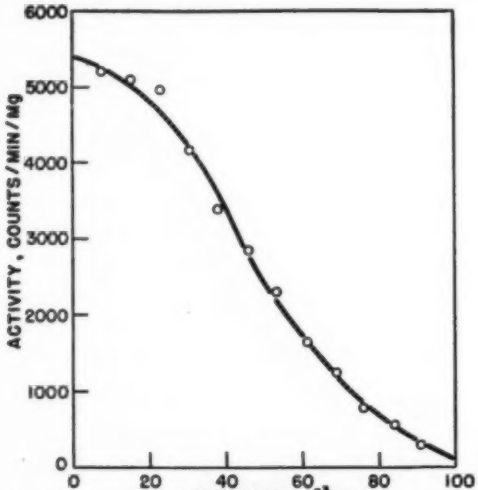


FIG. 2a. Typical activity-penetration curve for a single crystal at 870°C.

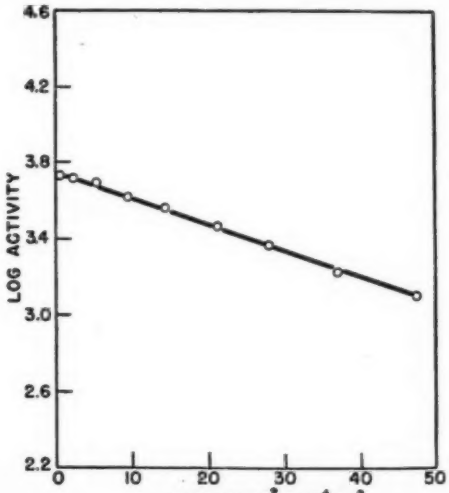


FIG. 2b. Same data as Fig. 2a, plotted as log activity *vs* (distance)².

ture section (±1°C) 4 inches in length at its center. Temperatures were calculated from the emf developed by a platinum-platinum-10 percent rhodium thermocouple wound onto the tube containing the specimen in such a way that the hot junction was adjacent to the specimen. The emf of the thermocouple was measured with a Leeds and Northrup Type K potentiometer. Readings were taken at various times during the course of the diffusion anneal, and it was found that the root-mean-square deviation from the mean absolute temperature was approximately 0.3 percent.

D. Determination of Diffusion Coefficients

Data suitable for calculation of self-diffusion coefficients were obtained by two methods. The sectioning method, described in detail elsewhere[11,14] was used when the penetration of the isotope was of the order of several mils. For smaller penetrations it is no longer possible to measure D with satisfactory accuracy by the sectioning technique. Extrapolation of high temperature data on the self-diffusion coefficients of silver indicated that in single crystals the mean penetration of isotope at 500° to 600°C ought to be of the order of a mil after annealing times of several weeks. In order to complete experiments on single crystals in this temperature range within a reasonable time, a more sensitive technique had to be used. The method chosen was that of Steigman, Shockley, and Nix,[1] wherein the count from the active face of the crystal is measured before and after the diffusion anneal. For constant geometry of the counting setup, and after decay of the isotope is corrected for, the number of counts from the active face should be less after diffusion anneal than before, owing to the absorption by inactive silver of radiation from active silver than has diffused into it.

The decrease of the surface activity after diffusion is determined by the change in the spatial distribution of the radioactive atoms; and this distribution in turn is determined uniquely, when diffusion is homogeneous, by the product $D_L t$, where D_L is the diffusion coefficient at the temperature of the experiment and t is the time. Formally,

$$I/I_0 = f(D_L t), \tag{4}$$

where I_0 is the initial, I the final, activity of the surface. Steigman, Shockley, and Nix expressed I/I_0 in terms of the absorption coefficient, μ, (measured independently of D_L) of the copper for the radiation. However, for silver, good values of D_L are already available at comparatively high temperatures from activity-penetration curves, and by measurement of I/I_0 as a function of time at a temperature where D_L is known, the functional relationship between I/I_0 and $D_L t$, valid also at any other temperature, can be determined. This procedure is particularly advantageous for silver because of the complicated nature of its absorption of radiation from Ag^{110}.

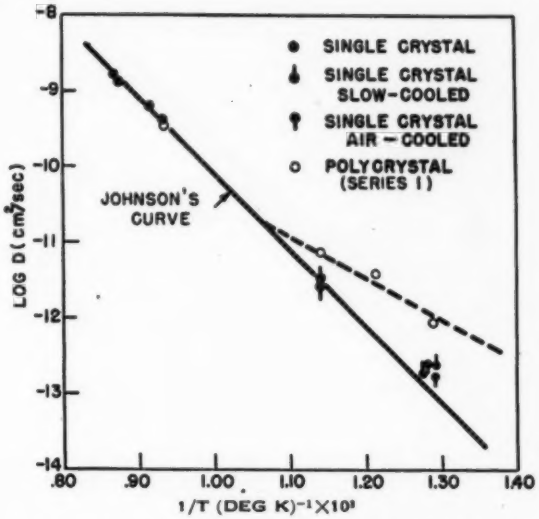


FIG. 3. Experimental results for single crystals and series 1 polycrystals.

In application of the method some important precautions must be observed:

1. Since the measured activity of a sample is very sensitive to the geometry of the system, the latter must be kept constant within narrow limits for all measurements that are to be intercompared.
2. The time-decay of the isotope and changing tube characteristics are most easily compensated for by use of a duplicate reference specimen which is counted at the same times that the experimental samples are counted, but is not subjected to the diffusion anneal. It is assumed that any change that occurs in the activity of the reference sample also occurs in proportionate amount in the activity of the experimental sample.
3. Experiments showed that I/I_0 is not sensibly different from unity for thicknesses, x, of silver absorbers as great as 0.2 mil but decreases rapidly for values of x exceeding 0.5 mil. It was found that interposition of a silver absorbing foil one mil thick between the specimen and the counter tube made it possible to determine I/I_0 as a function of $D_L t$ for small values of the latter with satisfactory accuracy.
4. To minimize statistical fluctuations a total of 20,000 to 40,000 counts should be taken on each specimen after any given treatment.
5. The thickness of the electrodeposited layer must be small (about 0.05 mil) in comparison with the mean penetration expected in the experiment.
6. Care must be taken to insure that none of the plated layer is lost mechanically from the surface during handling.

Figure 1 shows the measured curve of I/I_0 as a function of $D_L t$ for the geometrical and counting arrangement used in most of the experiments.

E. Counting Equipment

The counting arrangement consisted of a Tracerlab Autoscalar, Model SC1A, in connection with a Radiation Counter Laboratories Geiger-Müller tube, Model 10A. The tube and sample were mounted in a lead shield 2″ thick. The background amounted to 15–20

[14] P. H. Miller and F. Banks, Phys. Rev. **71**, 648 (1942).

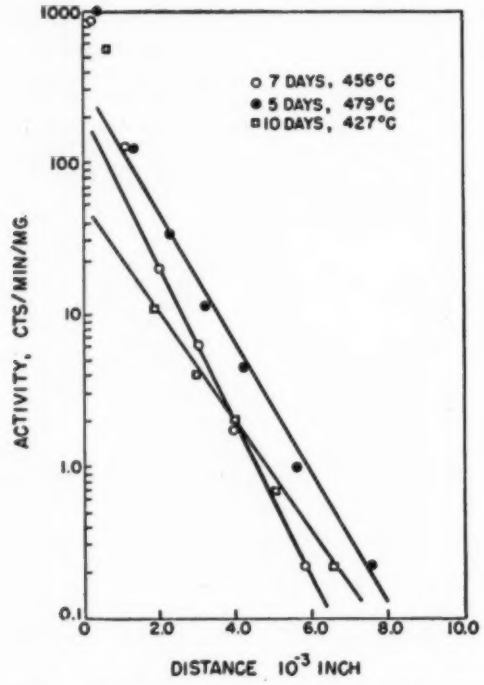


FIG. 4. Typical activity-penetration curves for polycrystals.

counts per minute. The total number of counts measured on any given sample varied from 1400 to 40,000, depending on the specific activity of the sample and the accuracy desired.

RESULTS AND DISCUSSION

A. Single Crystal and High Temperature Polycrystal Results

A typical activity-penetration curve obtained on a single crystal specimen is shown in Fig. 2a. Figure 2b shows that the same data can be represented as a linear relation between $\log a$ and y^2, as required by the theory of homogeneous diffusion:

$$\log a = \log[A/(D_L t)^{\frac{1}{2}}] - y^2/2.303 \cdot 4 \cdot D_L t, \quad (5)$$

where A is the total amount of activity in the deposited layer. Values of D_L were calculated from the measured slope of the $(\log a) - y^2$ curve. The results are shown in Fig. 3, wherein the experimental points are the results of this investigation and the solid line is drawn according to Johnson's relation.[11] In view of the excellent agreement between these two sets of data, it may be concluded that D_L calculated by Eq. (1) is the absolute lattice self-diffusion coefficient of silver.

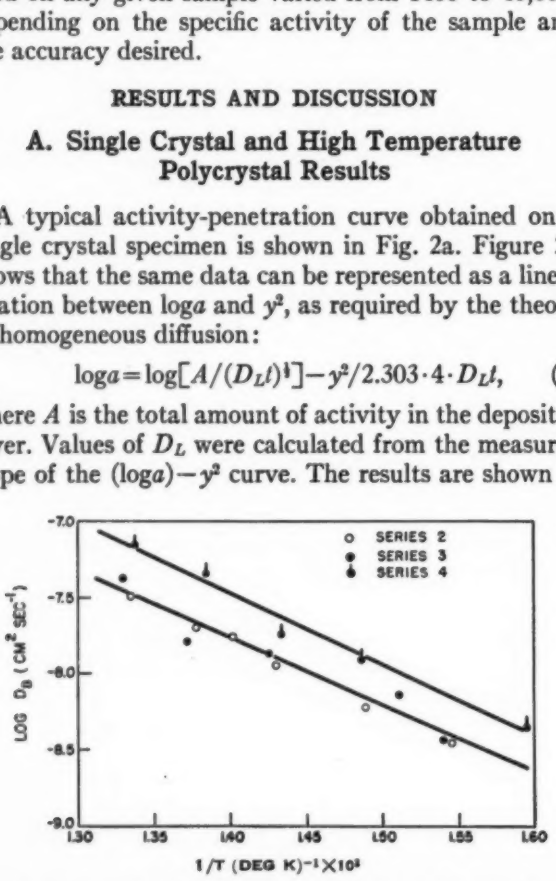


FIG. 5. Results of sectioning experiments on polycrystals.

The results of the experiments on the effect of thermal history on D_L are also shown in Fig. 3. Within the experimental error D_L is independent of the thermal histories used. The results can be explained on the basis of either of the following hypotheses:

1. The vacancies "leak" to the surface or some internal "precipitation point" of the supersaturated specimen in a time that is very small relative to the duration of the diffusion anneal.
2. The principal self-diffusion mechanism in silver does not involve lattice vacancies (unless associated with interstitial atoms).

B. Low Temperature Polycrystal Results

The preliminary measurements of the apparent diffusion coefficients in polycrystals (Series 1) in the temperature range 500–600°C as determined by the face-counting technique have been published previously[10] and discussed above. For purposes of comparison, these results are also shown graphically in Fig. 3.

TABLE II. Average grain sizes of specimens.

Series number	Mean linear dimension, cm
1	$3.6(10)^{-3}$
2	$6.0(10)^{-3}$
3	$2.1(10)^{-2}$
4	$4.6(10)^{-3}$

In order to obtain values of the absolute grain-boundary diffusion coefficient, D_B, activity-penetration curves were obtained at low temperatures and the results interpreted on the basis of Fisher's analysis.[12] Some typical penetration curves are shown in Fig. 4. As required by the analysis, the logarithm of the activity varies linearly with the depth of penetration, y, rather than with y^2 as in homogeneous lattice diffusion.

Values of D_B are calculated from Eq. (3) above, assuming δ to be 5×10^{-8} cm at all temperatures. With this assumption it is possible that the absolute magnitude of D_B may be in error by a factor of 2 or 3; but since δ probably is not a sensitive function of temperature, it is believed that the calculated activation energy is quite accurate.

The results of these measurements are shown in Fig. 5. When several measurements were made at the same temperature, only the average value is given. The straight lines connecting the points were calculated by the method of least squares. For the H-H silver, the equation of the straight line is

$$D_B = 0.03 \exp(-20{,}200/RT) \text{ cm}^2 \text{ sec}^{-1};$$

and for the J-M silver,

$$D_B = 0.12 \exp(-21{,}500/RT) \text{ cm}^2 \text{ sec}^{-1}.$$

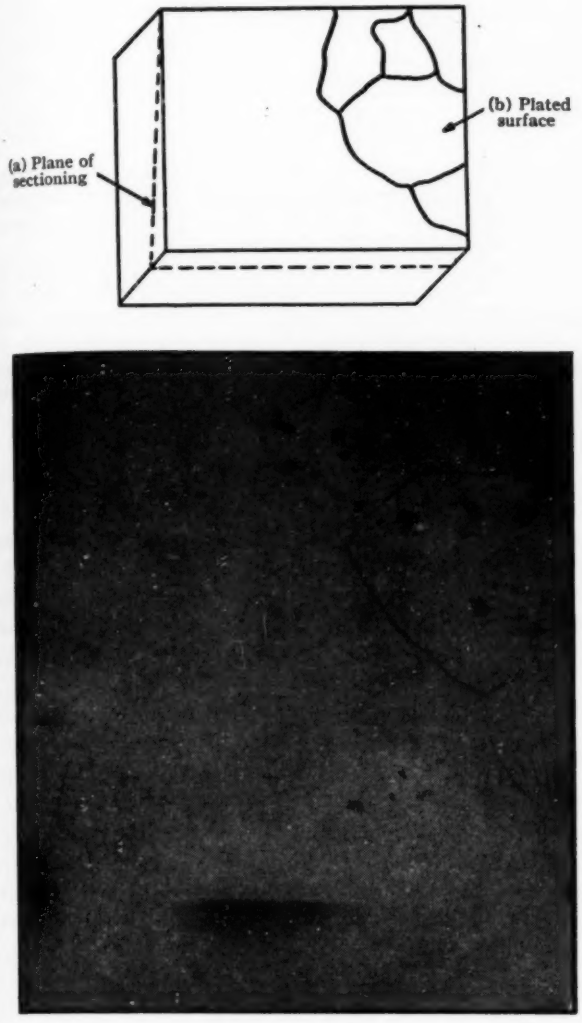


FIG. 6. Contact radiograph (b) of polycrystalline silver specimen sectioned as in (a) after a diffusion anneal.

The maximum deviation of an experimental point from that given by the straight line is 35 percent, the average deviation 20 percent.

The average grain sizes of the four series of specimens are given in Table II. These were measured on the part of the specimen remaining after sectioning. In no case was there an appreciable change in grain size during the diffusion. Furthermore, as is to be expected, D_B does not change as the grain size is altered in the same grade of silver (Fig. 5).

Birchenall and Mehl, on the basis of their results on self-diffusion in iron[15] and of Kê's interpretation of his anelasticity measurements,[16] have suggested that $D_B \approx D_L$ for self-diffusion in pure metals and that $D_B > D_L$ only when impurities are positively adsorbed at grain boundaries. It has been pointed out[17] that the grain sizes in their pure specimens may have been too large for their measured apparent diffusion coefficient, D_B', to be perceptibly different from D_L. On the other hand the results that are presented here indicate definitely that $D_B \gg D_L$ and that D_B does not depend significantly upon the variations in impurity content (0.03 percent to 0.001 percent nominal) of the silver.

Further confirmation that $D_B \gg D_L$ is shown in Fig. 6. A casting of H-H silver, machined to the shape sketched in Fig. 6a, was plated with radiosilver on the top surface. After a diffusion anneal of 4 hr at 450°C, the specimen was sectioned on the indicated plane, and the exposed section was put in contact with a film sensitive to radiation from the isotope. The dark lines of the resulting autoradiograph (see Fig. 6b) match perfectly with the grain boundaries of the specimen revealed by etching. It is clear that the isotope has penetrated at the grain boundaries much more deeply than into the crystal body. In view of the evidence, it seems quite certain that $D_B \gg D_L$ in silver self-diffusion.

It seems likely that D_B is a function of relative grain and boundary orientations. In the experiments described, the relative grain orientations were random, at least with respect to the axes perpendicular to the diffusion direction. In such systems most of the grain-boundary energies are relatively large, usually approaching the maximum possible due to orientation difference.[18] Experiments designed to measure D_B in silver as a function of relative crystal and boundary orientations are now in progress.

ACKNOWLEDGMENT

The authors are pleased to express their appreciation to Miss Mary Lou Feller (now Mrs. E. Epremian), Miss Betty Cleland, and to Mr. Robert Lewinski for their help in making the experimental measurements. Stimulating discussions with Dr. J. H. Hollomon and other members of the Metallurgy Division of the General Electric Research Laboratory are gratefully acknowledged.

[15] C. E. Birchenall and R. F. Mehl, Trans. Am. Inst. Mining Met. Engrs. **188**, 144 (1950).
[16] T. S. Kê, Phys. Rev. **73**, 267 (1948).
[17] R. E. Hoffman and D. Turnbull, Discussion to reference 15.
[18] C. S. Smith, Trans. Am. Inst. Mining Met. Engrs. **175**, 15 (1948).

On Shock-Wave Phenomena: Waterlike Substances

R. J. SEEGER AND H. POLACHEK
Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received November 17, 1950)

The properties of shock waves in waterlike substances are investigated in detail. A waterlike substance is defined as one for which the intrinsic energy is separable into a sum of two terms, one of which is a function of density only and the second a function of specific entropy. It is known that many substances, including water, behave in this fashion within a large range of pressure variation.

The paper consists of six sections. Section I contains some introductory material. Section II consists of a derivation of the basic equations that govern the propagation of plane shock waves in a waterlike substance. Section III is an analysis of one-dimensional interaction of shock waves, rarefaction waves, and contact discontinuities in a medium of this type. It is shown that a number of intrinsic differences exist between the behavior of shock waves in waterlike substances and their behavior in ideal gases.

In Section IV, one-dimensional interaction of shock waves in arbitrary fluids is discussed. Section V presents the theory of regular reflection of shock waves in waterlike substances; while Section VI presents the theory of triple-shock intersections. In both instances essential differences between the behavior of shock waves in waterlike substances and ideal gases are noted. The text is supplemented by numerous tables and charts.

I. INTRODUCTION

SHOCK-WAVE phenomena are familiar in all explosions. Much less attention, however, has been given to explosions in liquids than to those in gases.[1] The present discussion is concerned primarily with the influence of thermodynamic factors on shock-wave phenomena, particularly on the interaction of shock waves in substances that behave like water.

A waterlike substance[2] is defined as one for which the intrinsic energy E (per unit mass) is separable into one part E_1 that depends only upon the density ρ and a second part E_2 that depends only upon the specific entropy S, i.e.,

$$E(\rho, S)=E_1(\rho)+E_2(S).$$

The pressure p and the temperature T are given thermodynamically by

$$p=-\partial E/\partial(1/\rho),$$

and

$$T=\partial E/\partial S.$$

Hence, $p=p(\rho)$ under all conditions. For example, explosion processes in water are characterized over a range of high pressures by the relation $p=A\rho^n$ (A, n: constants).

In the case of polytropic changes (constant specific heat) in an ideal gas $p=B(S)\rho^{\gamma'}$, where γ' is the so-called polytropic exponent. Thus, $p=B\rho^{\gamma}$ (B, γ: constants) for an isentropic, polytropic change, where γ is the ratio of the specific heat at constant pressure to that at constant volume. A gas undergoing such a specialized set of changes, however, does not fulfill the general requirement for a waterlike substance.

The phenomenon of the hydraulic jump, on the other hand, for a liquid flowing in an open shallow channel is described exactly by the waterlike relation $p_h=A\rho_h^2$, where p_h and ρ_h are, respectively, the values of the pressure and of the density integrated over the depth. It has become customary to regard the hydraulic jump as an analogue of a weak shock in an ideal gas with $\gamma=2$, inasmuch as the Rankine-Hugoniot relation for weak shock waves reduces to $p=C\rho^{\gamma}$. The coefficient C, however, is a constant only up to third-order variations in the entropy. Strictly speaking, therefore, we describe a weak shock wave as a polytropic process and not as an analogue of the hydraulic jump. A critical examination of the mathematical limit of a weak shock indicates that the interactions of shock waves in waterlike substances exhibit, both qualitatively and quantitatively, different behaviors from those in ideal gases undergoing polytropic changes.

II. PLANE SHOCK WAVES IN WATERLIKE SUBSTANCES

Consider a simple model[1] of a shock wave, say, a plane step shock wave I (Fig. 1a) moving with constant velocity U in a liquid of negligible viscosity and thermal conductivity. Let the initial physical state of the liquid be specified by its pressure p_0, density ρ_0, and material velocity $u_0=0$ (for convenience), and that behind the shock be specified by pressure p, density ρ, and material velocity V (normal to the shock).

If we consider the flow of the liquid relative to the shock front (Fig. 1b), we have by the conservation of mass

$$\rho_0 U=\rho(U-V). \tag{1}$$

The conservation of momentum requires that

$$p-p_0=(\rho_0 U)V. \tag{2}$$

Eliminating V from Eqs. (1) and (2), we obtain for the speed of the shock wave relative to the medium ahead of it

$$U=[(p-p_0)/(\rho_0/\rho)(\rho-\rho_0)]^{\frac{1}{2}}. \tag{3a}$$

Substituting this value of U in Eq. (2), we have

$$V=[(p-p_0)(\rho-\rho_0)/\rho\rho_0]^{\frac{1}{2}}. \tag{4}$$

[1] H. Polachek and R. J. Seeger, "On shock-wave phenomena: interaction of shock waves in gases," Am. Math. Soc., Proceedings of Symposia in Applied Mathematics I (1949). *Regular Reflection of Shocks in Waterlike Substances* (Bureau of Ordnance, Explosives Research Report No. 14, 1944).

[2] J. von Neumann, *Oblique Reflection of Shocks* (Bureau of Ordnance, Explosives Research Report No. 12, 1943).

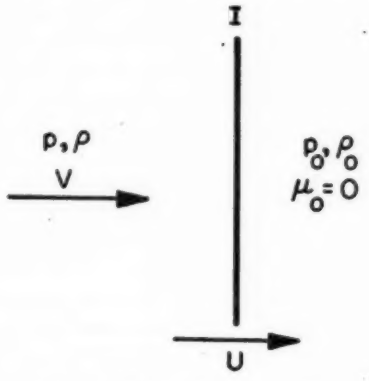


FIG. 1a. Plane step-shock wave.

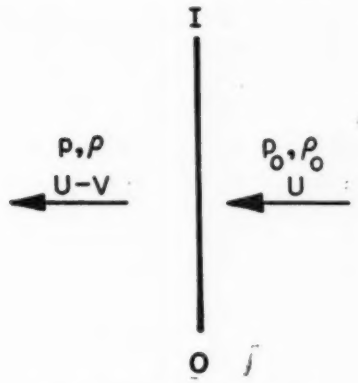


FIG. 1b. Flow of liquid relative to shock front.

Hence, the shock speed relative to the medium behind it is given by

$$U-V=[(p-p_0)/(\rho/\rho_0)(\rho-\rho_0)]^{\frac{1}{2}}. \tag{3b}$$

The conservation of energy states

$$E-E_0=(p+p_0)(\rho-\rho_0)/2\rho\rho_0. \tag{5}$$

Tammann's equation of state,[3] which is useful for some liquids at high pressures, is

$$(p+a)(v-b)=KT, \tag{6}$$

where $v(=1/\rho)$ is the specific volume and a, b, and K are constants for an individual liquid. In this case the change in intrinsic energy is given by

$$E-E_0=\bar{c}_v(T-T_0)+a(v-v_0), \tag{7a}$$

where $\bar{c}_v$ is an average specific heat for constant volume, i.e.,

$$\bar{c}_v=\int_{T_0}^{T} c_v dT/(T-T_0).$$

Eliminating T from Eqs. (6) and (7a) we obtain

$$E-E_0=(\bar{c}_v/K)\{(p+a)(v-b) -(p_0+a)(v_0-b)\}+a(v-v_0). \tag{7b}$$

The substitution of this expression in Eq. (5) gives

$$\frac{v-b}{v_0-b}=\frac{[(2\bar{c}_v/K)+1](p_0+a)/(p+a)+1}{[2\bar{c}_v/K+1]+(p_0+a)/(p+a)}.$$

Or, in terms of the density ratio $\eta\equiv(v-b)/(v_0-b)$ and of the compression ratio $\xi\equiv(p_0+a)/(p+a)$, we have the Rankine-Hugoniot dynamic adiabatic, i.e.,

$$\eta=[(\gamma+1)\xi+(\gamma-1)]/[(\gamma-1)\xi+(\gamma+1)], \tag{8a}$$

where $\gamma=1+(K/\bar{c}_v)$ reduces to the ordinary ratio of specific heats for an ideal gas ($a=b=0$, $K=R$).

It is possible to express all the previous relations in terms of a single parameter such as ξ. At the same time it is convenient to express all speeds as ratios with respect to some acoustic speed. Let

$$c_0\equiv[\gamma(p_0+a)v_0^2/(v_0-b)]^{\frac{1}{2}}$$

be the acoustic speed in the medium ahead of the shock, and $c\equiv[\gamma(p+a)v^2/(v-b)]^{\frac{1}{2}}$ be the acoustic speed in the medium behind the shock. Then Eqs. (3a) and (3b) can be expressed, respectively, as follows:

$$M\equiv\frac{U}{c_0}=\left[\frac{(\gamma+1)\xi^{-1}+(\gamma-1)}{2\gamma}\right]^{\frac{1}{2}}, \tag{9a}$$

$$\sigma\equiv\frac{U-V}{c}=\left[\frac{(\gamma+1)\xi+(\gamma-1)}{2\gamma}\right]^{\frac{1}{2}}. \tag{9b}$$

In the case of explosions in water, the constants take on the following average values in the domain of physical interest: $a=3000$ atmos, $b=0$, $\gamma=7.15$. Consequently, most shocks are comparatively weak, i.e., p_0, $p\ll a(\xi\sim 1)$, so that entropy changes are practically negligible. The dynamic Rankine-Hugoniot values (Fig. 2), based upon several different assumptions as to the

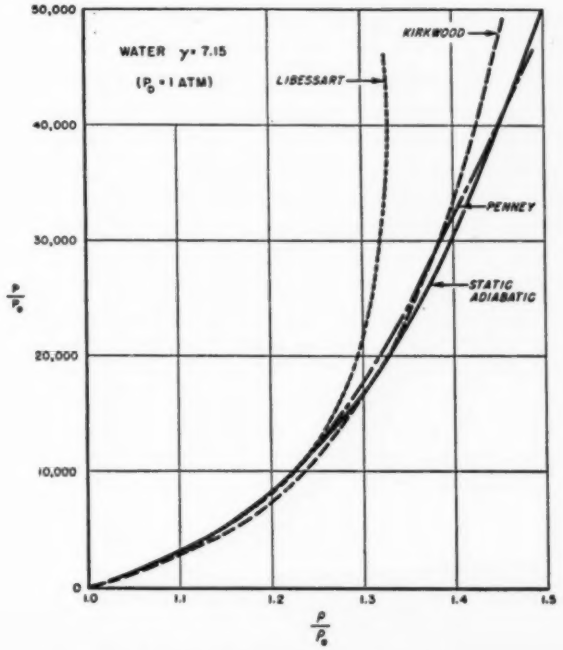


FIG. 2. Pressure density chart.

[3] G. Tammann, Ann. Physik 37, 975 (1912).

temperature variation of c_v, are approximated by the static adiabatic values. We assume that under all conditions

$$\eta=\xi^{1/\gamma}; \tag{8b}$$

hence, water is a special case of a waterlike substance.

In the case of the hydraulic jump, there is a similar equation with the effective γ having the value 2.

Two additional quantities can readily be expressed in terms of ξ whenever $b=0$, namely,

$$\tau\equiv\frac{U}{c}=\xi^{-1/\gamma}\left(\frac{1-\xi}{\gamma(\xi^{-1/\gamma}-1)}\right)^{\frac{1}{2}}, \tag{9c}$$

and

$$-v\equiv\frac{V}{c}=\pm\left(\frac{(1-\xi)(\xi^{-1/\gamma}-1)}{\gamma}\right)^{\frac{1}{2}}. \tag{9d}$$

Here τ represents the speed of the shock relative to the medium ahead; it differs from M in the referent acoustic speed. The quantity v signifies the change in the material speed across the shock front [the sign $-v$ corresponds to that of $(1-\xi)$]. In general,

$$\tau=\sigma-v,$$

and

$$\eta=\sigma/\tau.$$

The same formulas hold for normalization n relative to the region on the other side of the shock wave, provided only that $\xi=(p+a)/(p_n+a)$. Then σ is essentially the shock speed relative to the fluid in the normalized region, τ the shock speed relative to the region not normalized, v the increase of material speed across the shock relative to the normalized region [same sign as $(\xi-1)$], and c/c_n the acoustic speed in the region not normalized.

III. ONE-DIMENSIONAL INTERACTION FOR WATERLIKE SUBSTANCES

The one-dimensional interaction of shock waves, rarefaction waves, and contact-discontinuities (density alone discontinuous) can be conveniently discussed for waterlike substances in the manner used by Courant and Friedrichs for ideal gases.[4] One considers the change in material speed in terms of the characterizing pressure ratio ξ.

Consider a plane shock wave. Let V_n be the material speed of the normalized region and V the material speed of the other one. Then, we have from the definition of v that

$$V_n=V-c_nv(\xi).$$

Put

$$\Phi_n(\bar{p})=c_nv(\xi),$$

where

$$\bar{p}=p+a,$$

the effective pressure. Then

$$V=V_n+\Phi_n(\bar{p}). \tag{10a}$$

If the sense of the direction from left to right be always positive, then Eq. (10a) applies to Fig. 1a. Such a forward shock wave will be designated S_R. On the other hand, a backward shock, traveling toward the left, will be indicated by S_L and will satisfy the following relation:

$$V=V_n-\Phi_n(\bar{p}). \tag{10b}$$

The various shocks that are possible for a given value of V or of V_n may be conveniently represented on a V, $\bar{p}$ diagram (Fig. 3a and 3b). It is to be noted that if a point k is given on a V, $\bar{p}$ diagram, a single infinity of curves $V=V_k+\Phi_k(\bar{p})$ can be drawn through that point, since $\Phi_k(\bar{p})$ is a function of $\bar{p}$ and of ρ_k. If ρ_k, however, is also fixed, then the curve through k is determined.

Consider a finite rarefaction wave R_R (continuous adiabatic change) moving toward the right. As before, let V_n be the material speed of the normalized region and V be that of the region not normalized (i.e., on either side of the rarefaction). Then, we have from Riemann's argument[5]

$$V-(2/\gamma-1)c=V_n-(2/\gamma-1)c_n.$$

Put

$$\Psi_n(\bar{p})\equiv-(2c/\gamma-1)[(c_n/c)-1].$$

Then

$$V=V_n+\Psi_n(\bar{p}). \tag{11a}$$

In the case of such a rarefaction wave R_L traveling to the left

$$V=V_n-\Psi_n(\bar{p}). \tag{11b}$$

The V, $\bar{p}$ diagram for such rarefaction waves are shown in Figs. 3c and 3d.

The functions $\Phi_n(\bar{p})$ and $\Psi_n(\bar{p})$ represent physically the absolute change in material speed across a shock wave and a rarefaction wave of finite amplitude, respectively. They are connected by certain intrinsic relations, which we shall first list.

$$\Phi_n(\bar{p})=\Psi_n(\bar{p})=0, \qquad \bar{p}=\bar{p}_n. \tag{12a}$$

$$\Phi_n(\bar{p})\gtrless\Psi_n(\bar{p}), \qquad \bar{p}\gtrless\bar{p}_n. \tag{12a'}$$

$$\Phi_n'(\bar{p})=\Psi_n'(\bar{p})=c_n/\gamma, \qquad \bar{p}\to\bar{p}_n. \tag{12b}$$

$$\Phi_n'(\bar{p})\gtrless\Psi_n'(\bar{p}), \qquad \bar{p}\gtrless\bar{p}_n. \tag{12b'}$$

$$\Phi_n'(\bar{p}),\ \Psi_n'(\bar{p}) \text{ positive}, \qquad \text{all } \bar{p}. \tag{12b''}$$

$$\Phi_n''(\bar{p})=\Psi_n''(\bar{p})=-(1+\gamma)c_n/2\gamma^2, \qquad \bar{p}\to\bar{p}_n. \tag{12c}$$

$$\Phi_n''(\bar{p}),\ \Psi_n''(\bar{p}) \text{ negative}, \qquad \text{all } \bar{p}. \tag{12c'}$$

If $\rho_i\lesseqgtr\rho_j$ and $\bar{p}_i<\bar{p}_j$, then

$$\Phi_i(\bar{p})>\Phi_j(\bar{p}), \quad \bar{p}\gtrless\bar{p}_i. \tag{12d}$$

If

$$\bar{p}_i^{1/\gamma}/\rho_i\lesseqgtr\bar{p}_j^{1/\gamma}/\rho_j \quad \text{and} \quad \bar{p}_i<\bar{p}_j,$$

then

$$\Psi_i(\bar{p})>\Psi_j(\bar{p}), \quad \bar{p}\lesseqgtr\bar{p}_j. \tag{12d'}$$

$$\Phi_i(\bar{p}_k)>\Phi_i(\bar{p}_j)+\Phi_j(\bar{p}_k), \quad \bar{p}_i<\bar{p}_j<\bar{p}_k. \tag{12e}$$

$$\Psi_i(\bar{p}_k)=\Psi_i(\bar{p}_j)+\Psi_j(\bar{p}_k), \quad \text{all } \bar{p}_i, \bar{p}_j, \bar{p}_k. \tag{12e'}$$

[4] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948).

[5] B. Riemann, "Uber die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite," Gott. Abh. VIII, 43 (1858–59).

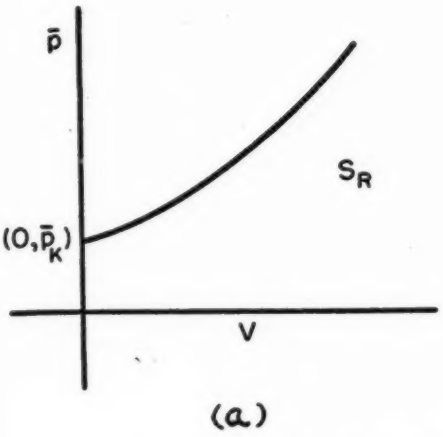


FIG. 3a. Velocity pressure diagram (forward shock wave).

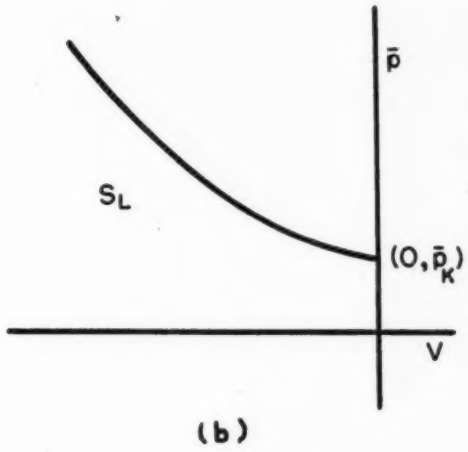


FIG. 3b. Velocity pressure diagram (backward shock wave).

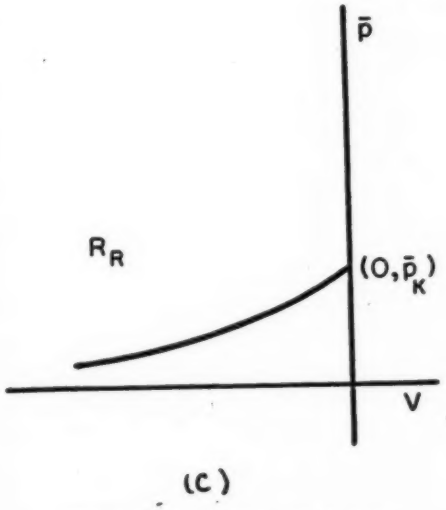


FIG. 3c. Velocity pressure diagram (forward rarefaction wave).

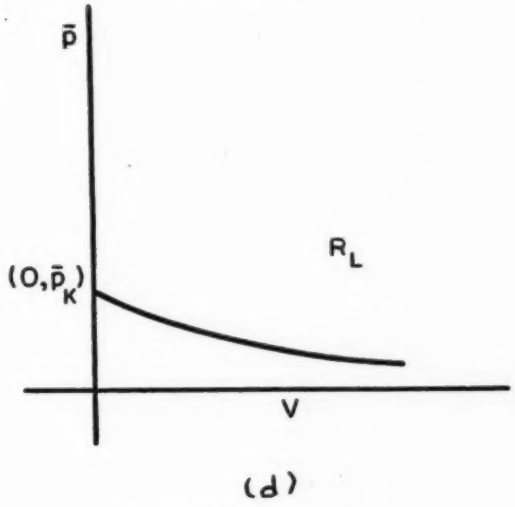


FIG. 3d. Velocity pressure diagram (backward rarefaction wave).

The proofs are based on the behavior of the functions $\Phi_n(\bar{p})$ and $\Psi_n(\bar{p})$ and their derivatives, which may be written as follows:

$$\Phi_n(\bar{p})=c_n v(\bar{p}/\bar{p}_n)=\pm(c_n/\gamma^{\frac{1}{2}})(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}, \quad [\pm, \text{ the sign of } (\xi-1)], \tag{13a}$$

$$\Phi_n'(\bar{p})=c_n v'\left(\frac{\bar{p}}{\bar{p}_n}\right)=\frac{c_n}{2\gamma^{\frac{1}{2}}\gamma}\frac{(\xi-1)\xi^{-(1+\gamma)/\gamma}+\gamma(1-\xi^{-1/\gamma})}{(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}}, \tag{13a'}$$

$$\Phi''(\bar{p})=c_n v''\left(\frac{\bar{p}}{\bar{p}_n}\right)=\frac{c_n}{2\gamma^2 v(\xi)}\left\{\frac{\gamma-1}{\gamma}\xi^{-(1+\gamma)/\gamma}+\frac{\gamma+1}{\gamma}\xi^{-(1+2\gamma)/\gamma}-2\gamma^2[v'(\xi)]^2\right\}. \tag{13a''}$$

Also,

$$\Psi_n(\bar{p})=2c_n/(\gamma-1)(\xi^{(\gamma-1)/2\gamma}-1), \tag{13b}$$

$$\Psi_n'(\bar{p})=(c_n/\gamma)\xi^{-(1+\gamma)/2\gamma}, \tag{13b'}$$

$$\Psi_n''(\bar{p})=-[(1+\gamma)c_n/2\gamma^2]\xi^{-(1+3\gamma)/2\gamma}. \tag{13b''}$$

1. Proof of Eqs. (12a′) and (12b′)

For any real nonzero value of A, B

$$A^2-2AB+B^2>0.$$

Let

$$A^2\equiv(\xi-1)\xi^{-(1+\gamma)/\gamma}, \quad B^2\equiv\gamma(1-\xi^{-1/\gamma}),$$

then,

$$(\xi-1)\xi^{-(1+\gamma)/\gamma}-2\gamma^{\frac{1}{2}}(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}\xi^{-(1+\gamma)/2\gamma}+\gamma(1-\xi^{-1/\gamma})>0,$$

or

$$\{(\xi-1)\xi^{-(1+\gamma)/\gamma}+\gamma(1-\xi^{-1/\gamma})\}c_n>2\gamma\gamma^{\frac{1}{2}}(\xi-1)^{\frac{1}{2}}\times(1-\xi^{-1/\gamma})^{\frac{1}{2}}\xi^{-(1+\gamma)/2\gamma}c_n/\gamma.$$

(a) For values of $0<\xi<1$; the expression

$$2\gamma\gamma^{\frac{1}{2}}(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}$$

is negative by the rule of signs of Eq. (13a); hence,

$$\frac{\{(\xi-1)\xi^{-(1+\gamma)/\gamma}+\gamma(1-\xi^{-1/\gamma})\}c_n}{2\gamma\gamma^{\frac{1}{2}}(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}}<\frac{\xi^{-(1+\gamma)/2\gamma}c_n}{\gamma},$$

or

$$\Phi_n'(\bar{p})<\Psi_n'(\bar{p}).$$

(b) For values of $\infty>\xi>1$; the expression

$$2\gamma\gamma^{\frac{1}{2}}(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}$$

is positive, hence,

$$\frac{(\xi-1)\xi^{-(1+\gamma)/\gamma}+\gamma(1-\xi^{-1/\gamma})}{(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}}\frac{c_n}{2\gamma\gamma^{\frac{1}{2}}}>\frac{\xi^{-(1+\gamma)/2\gamma}}{\gamma}c_n,$$

or

$$\Phi_n'(\bar{p})>\Psi_n'(\bar{p}).$$

Since $\Phi_n(\bar{p})=\Psi_n(\bar{p})$ and $\Phi_n'(\bar{p})=\Psi_n'(\bar{p})$ for $\bar{p}\to\bar{p}_n$ [Eqs. (12a) and (12b)], it follows that $\Phi_n(\bar{p})\gtrless\Psi_n(\bar{p})$ for $\bar{p}\gtrless\bar{p}_n$.

2. Proof of Eqs. (12a), (12b), (12c), (12b″), (12c″)

As $\bar{p}\to\bar{p}_n$, $\xi\to1$. It is evident that $\Phi_n(\bar{p})=\Psi_n(\bar{p})=0$ for $\xi\to1$. It also follows from Eqs. (13a′) and (13b′) that $\Phi_n'(\bar{p})=\Psi_n'(\bar{p})\to c_n/\gamma$, and from Eqs. (13a″) and (13b″) that $\Phi_n''(\bar{p})=\Psi_n''(\bar{p})\to-(1+\gamma)c_n/2\gamma^2$. From Eqs. (13b′) and (13b″) it follows that $\Psi_n'(\bar{p})>0$ and that $\Psi_n(\bar{p})<0$. It remains to be proven that $\Phi_n'(\bar{p})>0$ and that $\Phi_n''(\bar{p})<0$ for all values of $\bar{p}$. We shall first prove the former statement.

(a) For values of $0<\xi<1$: the expression

$$\Phi_n'(\bar{p})=\frac{c_n}{2\gamma\gamma^{\frac{1}{2}}}\frac{(\xi-1)\xi^{-(1+\gamma)/\gamma}+\gamma(1-\xi^{-1/\gamma})}{(\xi-1)^{\frac{1}{2}}(1-\xi^{-1/\gamma})^{\frac{1}{2}}}$$

is always positive, since the numerator is always negative and the denominator is always negative, by the rule of signs of Eq. (13).

(b) For $\xi\to1$: we have noted that $\Phi_n'(\bar{p})\to c_n/\gamma$, which is positive.

(c) For values of $\infty>\xi>1$: it is sufficient to show that $\Phi_n'(\bar{p})>\Psi_n'(\bar{p})$, since we have noted that $\Psi_n'(\bar{p})$ is always greater than zero. This has already been shown.

We will now prove that $\Phi_n''(\bar{p})<0$ for all values of $\bar{p}$. Now

$$\Phi_n''(\bar{p})=\frac{c_n}{2\gamma^2v(\xi)}\left\{\frac{\gamma-1}{\gamma}\xi^{-(1+\gamma)/\gamma}+\frac{\gamma+1}{\gamma}\xi^{-(1+\gamma)/\gamma}\xi^{-1}-2\gamma^2[v'(\xi)]^2\right\}.$$

(a′) For values of $0<\xi<1$: (1) v is negative from Eqs. (13a), and (2) $v'(\xi)<(1/\gamma)\xi^{-(1+\gamma)/2\gamma}$ or $\gamma^2v'^2<\xi^{-(1+\gamma)/\gamma}$ from the preceding section. To prove that $\Phi_n''(\bar{p})<0$, it remains to be shown that

$$2\gamma^2[v'(\xi)]^2<\xi^{-(1+\gamma)/\gamma}\{[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)\},$$

or

$$2\xi^{-(1+\gamma)/\gamma}<\xi^{-(1+\gamma)/\gamma}\{[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)\},$$

or $2<[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)$, which is true.

(b′) For $\xi\to1$: $\Phi_n''(\bar{p})\to-(1+\gamma)c_n/2\gamma^2$, which is negative.

(c′) For values of $\infty>\xi>1$: (1) v is positive from Eq. (13a), (2) $v'(\xi)>(1/\gamma)\xi^{-(\gamma+1)/2\gamma}$, or $\gamma^2[v'(\xi)]^2>\xi^{-(\gamma+1)/\gamma}$. To prove that $\Phi_n''(\bar{p})<0$ it remains to be shown that

$$2\gamma^2[v'(\xi)]^2>\xi^{-(\gamma+1)/\gamma}\{[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)\},$$

or

$$2\xi^{-(\gamma+1)/\gamma}>\xi^{-(\gamma+1)/\gamma}\{[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)\},$$

or $2>\{[(\gamma-1)/\gamma]+[(\gamma+1)/\gamma](1/\xi)\}$, which is true.

3. Proof of Eq. (12d)

We have

$$\Phi_i(\bar{p})=c_iv\left(\frac{\bar{p}}{\bar{p}_i}\right)=\left[\frac{\bar{p}_i}{\rho_i}\left(\frac{\bar{p}}{\bar{p}_i}-1\right)\left(1-\left(\frac{\bar{p}}{\bar{p}_i}\right)^{-1/\gamma}\right)\right]^{\frac{1}{2}}$$

$$=\left[\frac{1}{\rho_i}(\bar{p}-\bar{p}_i)\left(1-\left(\frac{\bar{p}}{\bar{p}_i}\right)^{-1/\gamma}\right)\right]^{\frac{1}{2}},$$

and

$$\Phi_j(\bar{p})=c_jv\left(\frac{\bar{p}}{\bar{p}_j}\right)=\left[\frac{\bar{p}_j}{\rho_j}\left(\frac{\bar{p}}{\bar{p}_j}-1\right)\left(1-\left(\frac{\bar{p}}{\bar{p}_j}\right)^{-1/\gamma}\right)\right]^{\frac{1}{2}}$$

$$=\left[\frac{1}{\rho_j}(\bar{p}-\bar{p}_j)\left(1-\left(\frac{\bar{p}}{\bar{p}_j}\right)^{-1/\gamma}\right)\right]^{\frac{1}{2}}.$$

If $\bar{p}=\bar{p}_i$, $\Phi_i(\bar{p})=0$ and $\Phi_j(\bar{p})$ is negative, then theorem (12d) is true. If $\bar{p}=\bar{p}_j$, $\Phi_i(\bar{p})$ is positive and $\Phi_j(\bar{p})=0$; if $\bar{p}_i<\bar{p}<\bar{p}_j$, $\Phi_i(\bar{p})$ is positive, while $\Phi_j(\bar{p})$ is negative; and, finally, if $\bar{p}_i<\bar{p}_j<\bar{p}$, both $\Phi_i(\bar{p})$ and $\Phi_j(\bar{p})$ are positive. In this case, however, the factors for the expression $\Phi_i(\bar{p})$ are each in absolute value either greater than or equal to the corresponding factors of $\Phi_j(\bar{p})$. Hence, in each case $\Phi_i(\bar{p})>\Phi_j(\bar{p})$, algebraically.

4. Proof of Eq. (12d′)

We have

$$\Psi_i(\bar{p})=\frac{2c_i}{\gamma-1}\left\{\left(\frac{\bar{p}}{\bar{p}_i}\right)^{(\gamma-1)/2\gamma}-1\right\}$$

$$=\frac{2\gamma^{\frac{1}{2}}}{\gamma-1}\left(\frac{\bar{p}_i^{1/\gamma}}{\rho_i}\right)^{\frac{1}{2}}\{\bar{p}^{(\gamma-1)/2\gamma}-\bar{p}_i^{(\gamma-1)/2\gamma}\},$$

and

$$\Psi_j(\bar{p})=\frac{2c_j}{\gamma-1}\left\{\left(\frac{\bar{p}}{\bar{p}_j}\right)^{(\gamma-1)/2\gamma}-1\right\}$$

$$=\frac{2\gamma^{\frac{1}{2}}}{\gamma-1}\left(\frac{\bar{p}_j^{1/\gamma}}{\rho_j}\right)^{\frac{1}{2}}\{\bar{p}^{(\gamma-1)/2\gamma}-\bar{p}_j^{(\gamma-1)/2\gamma}\}.$$

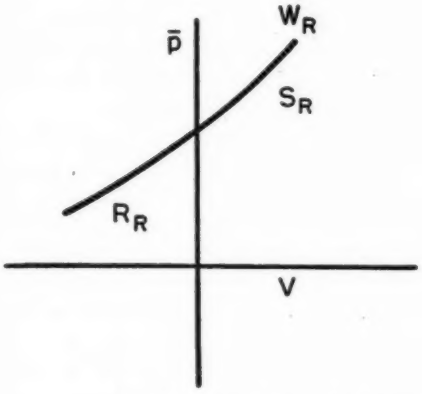


FIG. 4a. Forward transition curve.

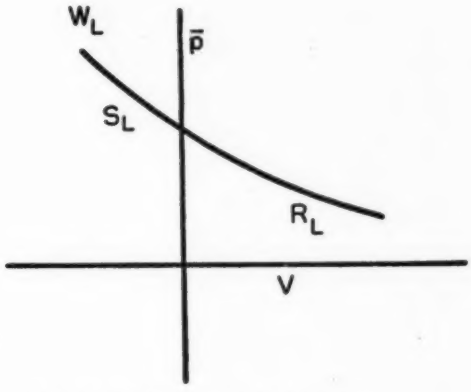


FIG. 4b. Backward transition curve.

If $\bar{p}=\bar{p}_i$, $\Psi_i(\bar{p})=0$ and $\Psi_j(\bar{p})$ is negative, then theorem (8d′) is satisfied. If $\bar{p}=\bar{p}_j$, $\Psi_i(\bar{p})$ is positive, while $\Psi_j(\bar{p})=0$: if $\bar{p}_i<\bar{p}<\bar{p}_j$, $\Psi_i(\bar{p})$ is positive and $\Psi_j(\bar{p})$ is negative: and, finally, if $\bar{p}<\bar{p}_i<\bar{p}_j$, both $\Psi_i(\bar{p})$ and $\Psi_j(\bar{p})$ are negative. In this case, however, the factors for the expression $\Psi_i(\bar{p})$ are each in absolute value either less than or equal to the corresponding factors of $\Psi_j(\bar{p})$. Hence, in each case $\Psi_i(\bar{p})>\Psi_j(\bar{p})$, algebraically.

5. Proof of Eq. (12e)

We are to prove that

$$v\left(\frac{\bar{p}_k}{\bar{p}_i}\right)c_i>v\left(\frac{\bar{p}_j}{\bar{p}_i}\right)c_i+v\left(\frac{\bar{p}_k}{\bar{p}_j}\right)c_j,$$

or

$$v\left(\frac{\bar{p}_k}{\bar{p}_j}\div\frac{\bar{p}_i}{\bar{p}_j}\right)\frac{c_i}{c_j}>v\left(\frac{1}{\bar{p}_i/\bar{p}_j}\right)\frac{c_i}{c_j}+v\left(\frac{\bar{p}_k}{\bar{p}_j}\right).$$

Let $\xi\equiv\bar{p}_i/\bar{p}_j$ and $\xi'\equiv\bar{p}_k/\bar{p}_j$, then $0<\xi<1<\xi'<\infty$. Since $c_i/c_j=(\bar{p}_i/\bar{p}_j)^{(\gamma-1)/2\gamma}=\xi^{(\gamma-1)/2\gamma}$, the above is equivalent to

$$v(\xi'/\xi)\xi^{(\gamma-1)/2\gamma}>v(1/\xi)\xi^{(\gamma-1)/2\gamma}+v(\xi').$$

Substituting the values of v, we obtain

$$\{(\xi'-\xi)(\xi^{-1/\gamma}-\xi'^{-1/\gamma})\}^{\frac{1}{2}}>\{(1-\xi)(\xi^{-1/\gamma}-1\}^{\frac{1}{2}}+\{(\xi'-1)(1-\xi'^{-1/\gamma})\}^{\frac{1}{2}}.$$

This relation may be proved by a method similar to that used in the proof of Eqs. (12a′) and (12b′). For any real nonzero value of A, B

$$A^2-2AB+B^2>0.$$

Let

$$A^2\equiv(\xi'-1)(\xi^{-1/\gamma}-1),\quad B^2\equiv(1-\xi)(1-\xi'^{-1/\gamma}),$$

then

$$(\xi'-1)(\xi^{-1/\gamma}-1)-2\{(1-\xi)(1-\xi'^{-1/\gamma})\}^{\frac{1}{2}}[(\xi'-1)(\xi^{-1/\gamma}-1)\}^{\frac{1}{2}}+(1-\xi)(1-\xi'^{-1/\gamma})>0,$$

or

$$(\xi'-1)(\xi^{-1/\gamma}-1)+(1-\xi)(1-\xi'^{-1/\gamma})+(\xi'-1)(1-\xi'^{-1/\gamma})+(1-\xi)(\xi^{-1/\gamma}-1)>(1-\xi)(\xi^{-1/\gamma}-1)+(\xi'-1)(1-\xi'^{-1/\gamma})+2\{(1-\xi)(\xi^{-1/\gamma}-1)\}^{\frac{1}{2}}\times\{(\xi'-1)(1-\xi'^{-1/\gamma})\}^{\frac{1}{2}},$$

or

$$\{(\xi'-1)+(1-\xi)\}\times\{(\xi^{-1/\gamma}-1)+(1-\xi'^{-1/\gamma})\}>(1-\xi)(\xi^{-1/\gamma}-1)+(\xi'-1)(1-\xi'^{-1/\gamma})+2\{(1-\xi)(\xi^{-1/\gamma}-1)\}^{\frac{1}{2}}\times\{(\xi'-1)(1-\xi'^{-1/\gamma})\}^{\frac{1}{2}}.$$

Hence,

$$(\xi'-\xi)(\xi^{-1/\gamma}-\xi'^{-1/\gamma})>\{[(1-\xi)(\xi^{-1/\gamma}-1)]^{\frac{1}{2}}+[(\xi'-1)(1-\xi'^{-1/\gamma})]^{\frac{1}{2}}\}^2.$$

Taking the square root of each side we obtain

$$\{(\xi'-\xi)(\xi^{-1/\gamma}-\xi'^{-1/\gamma})\}^{\frac{1}{2}}>\{[(1-\xi)(\xi^{-1/\gamma}-1)]^{\frac{1}{2}}+[(\xi'-1)(1-\xi'^{-1/\gamma})]^{\frac{1}{2}}\},$$

which is what we set out to prove.

6. Proof of Eq. (12e′)

This relation may be seen to be true by considering three points on a forward rarefaction wave $x_i<x_j<x_k$ with effective pressures $\bar{p}_i$, $\bar{p}_j$, and $\bar{p}_k$ and material speeds V_i, V_j, and V_k. The theorem simply states that

$$(V_k-V_i)=(V_j-V_i)+(V_k-V_j).$$

It is useful in discussing interaction on the basis of these theorems to construct for any given state a W_R transition curve for waves toward the right (Fig. 4a) and a W_L transition curve for those to the left (Fig. 4b).

For example, let us now consider the head-on collision of two shocks S_1 and S_2. It is required to determine the final state of the intermediary region, given its initial state (p_0, ρ_0, V_0) and the strengths ξ_1 and ξ_2 (for definiteness ξ_1 is taken less than ξ_2). Evidently (see the illustrative example for water in Fig. 5) a common V_0' and p_0' exists only for the intersection of the upper branches of the W transition curves through the initial V, $\bar{p}$ points. In other words, the original colliding shocks result in

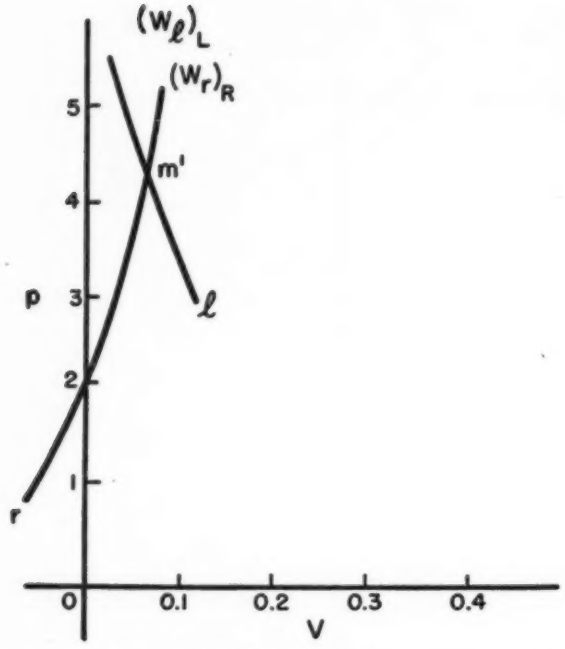


FIG. 5. Head-on collision of shock waves, illustrative problem.

$(S_1)_R(S_2)_L \rightarrow (S_3)_L(S_4)_R$

$(S_1)_R$	Before	$(S_2)_L$
$p_1=3.002$	$p_m=0.001$	$p_r=1.001$
$c_1=1.347$	$c_m=1.000$	$c_r=1.131$
$V_1=0.114_R$	$V_m=0.000$	$V_r=0.43_L$
$(S_3)_L$	After	$(S_4)_R$
$p_1=3.002$	$p_{m'}=4.450$	$p_r=1.001$
$c_1=1.347$	$c_{m'}=1.478$	$c_r=1.131$
$V_1=0.114_R$	$V_{m'}=0.0708_R$	$V_r=0.043_L$

two new shocks receding from each other. Thus, symbolically we may write

$$(S_1)_R(S_2)_L \rightarrow (S_3)_L(S_4)_R.$$

It is to be noted that because of the assumption of negligible entropy change, the condition of pressure equality between two regions implies density equality also, i.e., no contact discontinuity can exist in this case.[3] In a similar fashion other types of interaction can be discussed; they are summarized in Table I, which includes also the ideal gas results[3] for comparison. It is significant that a waterlike substance does not always behave like an ideal gas (fictitious for $\gamma>5/3$) having the same value of γ. For example, in the case of water the overtaking of one shock wave by another results always in a rarefaction being reflected, whereas there may be a reflected shock for the analogous ideal gas. Then, too, for a rarefaction overtaking a shock wave the reflected wave is always a shock in water, but it may be a rarefaction in the ideal gas. Moreover, the kinds of phenomena that occur in waterlike substances are independent of the value of γ, but not so in the gamut of so-called ideal gases, where $\gamma=5/3$ is a peculiar mathematical boundary (unrelated to the physical limitation). We shall now discuss in detail two instances of wave interactions, in which the differences between ideal gases and waterlike substances are most striking, *viz.*, a shock wave overtaking a shock and a rarefaction overtaking a shock wave.

We shall first recall some relations existing between the effective pressure, material speed, and density ahead and behind a shock wave or a rarefaction wave. Consider the shock or rarefaction wave moving, say, along the x-axis. The x-axis will, thus, be divided in two parts. The effective pressure, material speed, and density on the right of the wave front will be designated by $\bar{p}_r$, V_r, and ρ_r, respectively; those on the left, $\bar{p}_l$, V_l, and ρ_l, respectively. The material speed will be considered positive if it is moving in a positive direction along this x-axis, and negative if it is moving in the negative direction. A shock or rarefaction wave moving in a forward, or positive, direction relative to the undisturbed material will be designated S_R and R_R, respectively; in the opposite direction, S_L and R_L. On the basis of the above conventions, the following inequalities hold.

$$\begin{aligned} &\bar{p}_l>\bar{p}_r, \quad \rho_l>\rho_r, \quad \text{for} \quad S_R; \\ &\bar{p}_l<\bar{p}_r, \quad \rho_l<\rho_r, \quad \text{for} \quad S_L; \\ &V_l>V_r, \qquad\qquad \text{for} \quad S_R \text{ or } S_L. \end{aligned} \tag{A}$$

$$\begin{aligned} &\bar{p}_l<\bar{p}_r, \quad \rho_l<\rho_r, \quad \text{for} \quad R_R, \\ &\bar{p}_l>\bar{p}_r, \quad \rho_l>\rho_r, \quad \text{for} \quad R_L, \\ &V_l<V_r, \qquad\qquad \text{for} \quad R_R \text{ or } R_L. \end{aligned} \tag{A'}$$

From relations (12b″) and (12c′), it follows that $\Phi_n(\bar{p})$ and $\Psi_n(\bar{p})$ are monotonically increasing functions of $\bar{p}$, while their slopes are monotonically decreasing functions of $\bar{p}$. Before proceeding, we call attention to the relations $\Phi_r(\bar{p}_l)=-\Phi_l(\bar{p}_r)$ and $\Psi_r(\bar{p}_l)=-\Psi_l(\bar{p}_r)$

TABLE I.[a]

Waterlike substances	Ideal gas[b]
$S_R\,T_L \rightarrow S_L\,T\;\; S_R$	$S_R\,T_L \rightarrow S_L\,T\;\; S_R$
$S_R\,T_R \rightarrow R_L\,T\;\; S_R$	$S_R\,T_R \rightarrow R_L\,T\;\; S_R$
$R_R\,T_L \rightarrow R_L\,T\;\; R_R$	$R_R\,T_L \rightarrow R_L\,T\;\; R_R$
$R_R\,T_R \rightarrow S_L\,T\;\; T\;\; R_R$	$R_R\,T_R \rightarrow S_L\,T\;\; T\;\; R_R$
$S_R\,S_L \rightarrow S_L\,S_R$	$S_R\,S_L \rightarrow S_L\,T\;\; S_R$
$S_R\,R_L \rightarrow R_L\,S_R$	$S_R\,R_L \rightarrow R_L\,T\;\; T\;\; S_R$
$R_R\,R_L \rightarrow R_L\,R_R$	$R_R\,R_L \rightarrow R_L\,R_R$
$S_R\,S_L \rightarrow R_L\,S_R$	$S_R\,S_R \rightarrow R_L\,T\;\; S_R$ all γ
	$S_R\,S_R \rightarrow S_L\,T\;\; S_R$ $\}\gamma>5/3$
	$S_R\,S_R \rightarrow \;\; T\;\; S_R$ $\}\gamma>5/3$
$R_R\,S_R \rightarrow S_L\,R_R$	$R_R\,S_R \rightarrow S_L\,T\;\; T\;\; R_R$ $\}$
$R_R\,S_R \rightarrow S_L$	$R_R\,S_R \rightarrow S_L\,T\;\; T\;\; S_R$ $\}$
$R_R\,S_R \rightarrow S_L\,S_R$	$R_R\,S_R \rightarrow R_L\,T\;\; T\;\; S_R$ $\}$
	$R_R\,S_R \rightarrow R_L\,T\;\; T\;\; R_R$ $\}\gamma\leqslant 5/3$
	$R_R\,S_R \rightarrow S_L\,T\;\; T$ $\}$
	$R_R\,S_R \rightarrow \;\; T\;\; T$ $\}$
	$R_R\,S_R \rightarrow R_L\,T\;\; T$ $\}$
$S_R\,R_R \rightarrow S_L\,S_R$	$S_R\,R_R \rightarrow S_L\,T\;\; T\;\; S_R$ $\}$
$S_R\,R_R \rightarrow S_L$	$S_R\,R_R \rightarrow S_L\,T\;\; T$ $\}\gamma\leqslant 5/3$
$S_R\,R_R \rightarrow S_L\,R_R$	$S_R\,R_R \rightarrow S_L\,T\;\; T\;\; R_R$ $\}$

[a] T_L signifies a density discontinuity with the density of the region on the left less than that on the right. $T\;\;T$ signifies a contact zone of constant pressure throughout where there is a monotonic change of density.
[b] See reference 3.

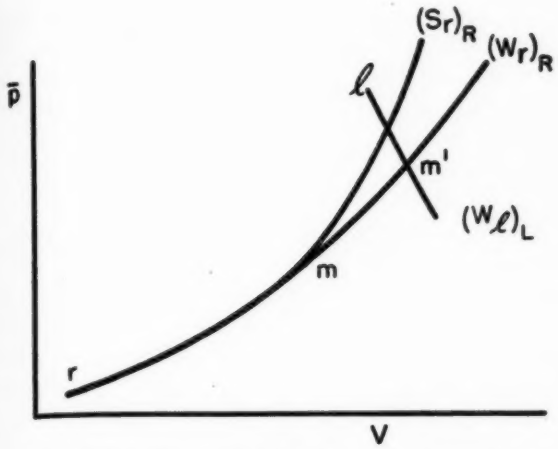


FIG. 6. Velocity pressure diagram for the interaction of two forward shocks.

for a shock wave and for a rarefaction wave, respectively, each of which simply states that the difference of material speed $(V_l - V_r) = -(V_r - V_l)$.

(a) $S_R S_R$. Let us consider the case of two forward shocks following each other. These will divide the x-axis into three regions, left, middle, and right, which we shall designate l, m, and r, respectively. From inequalities (A) we have the inequalities $\bar{p}_l > \bar{p}_m > \bar{p}_r$ and $V_l > V_m > V_r$, which will determine the relative positions of l, m, and r on the V, $\bar{p}$ diagram. After the interaction has taken place, the regions l and r remain unchanged, while a new region m' appears at the center to take the place of m. The region m' is connected to region l by a backward wave W_L and to region r by a forward wave W_R. If we draw a $(W_l)_L$ curve (Fig. 6) through l and a $(W_r)_R$ curve through r, m' will be determined by the intersection of these two curves. We wish to show that m' is on the shock branch of $(W_r)_R$, but on the rarefaction-wave portion of $(W_l)_L$.

Since m was connected to r by a shock, the curve $(W_r)_R$ passes through the point m. From the disposition of the point l and the slope of $(W_l)_L$, it is evident that m' is on the shock branch of $(W_r)_R$. Hence, the forward wave connecting m' with the region r will be a shock. We shall now draw $(S_m)_R$ through l and m. It follows from relations (12e) that $(S_m)_R$ will lie on the left of $(W_r)_R$ for values of $\bar{p} > \bar{p}_m$. Hence, m' will lie on the rarefaction branch of $(W_l)_L$.

(b) $R_R S_R$. Let us consider a forward shock overtaken by a forward rarefaction wave. The inequalities (A) and (A′) yield the following relations for the effective pressures and material speeds in regions l, m, and r: $\bar{p}_l < \bar{p}_m > \bar{p}_r$ and $V_l < V_m > V_r$. The points l and r will lie on the $(R_m)_R$ and $(S_m)_R$ curves, respectively (Fig. 7a). Now for $\bar{p} < \bar{p}_m$, $\Phi_m(\bar{p}) < \Psi_m(\bar{p})$ Eq. (12a′). Hence, the curve $(S_m)_R$ will lie above $(R_m)_R$. It may be shown that $(W_r)_R$ will also lie above $(R_m)_R$ for values of $\bar{p}$ below $\bar{p}_m$. It follows that the point m' will fall on the $(S_l)_L$ branch of the $(W_l)_R$ curve, resulting in a weak reflected shock. On the other hand, m' may fall either on the $(S_r)_R$ or $(R_r)_R$ branch of $(W_r)_R$ curve depending on the original position of l on the $(R_m)_R$ curve (Fig. 7b). Thus, for a strong shock and a weak rarefaction wave a shock will be transmitted, while the reverse is true for a weak shock and a strong rarefaction wave. It remains to be shown that $(W_r)_R$ will lie above $(R_m)_R$ for values of $\bar{p}$ below $\bar{p}_m$. The transition curve $(W_r)_R$ consists of two branches. It is composed of $(S_r)_R$ from r to m and of $(R_r)_R$ for values of $\bar{p}$ less than $\bar{p}_r$. It follows from Eq. (12e) that $(S_r)_R$ must lie above $(S_m)_R$ from r to m; hence, this portion is also above $(R_m)_R$. Also, if we draw a vertical line through r, we pass through some point t on $(R_m)_R$. Relation (12e′) implies that $(R_t)_R$ coincides with $(R_m)_R$. But $\bar{p}_t < \bar{p}_r$ and $\bar{p}_t^{1/\gamma}/\rho_t = \bar{p}_r^{1/\gamma}/\rho_r$; hence, it follows from relations (12d′) that $\Psi_t(\bar{p}) > \Psi_r(\bar{p})$ for $\bar{p} \leqslant \bar{p}_r$. Consequently, $(R_r)_R$ lies above $(R_t)_R \equiv (R_m)_R$.

IV. ONE-DIMENSIONAL INTERACTION IN ARBITRARY FLUIDS

A number of the properties exhibited in the interaction of shock waves and/or rarefaction waves are not peculiar to ideal gases or to waterlike substances but

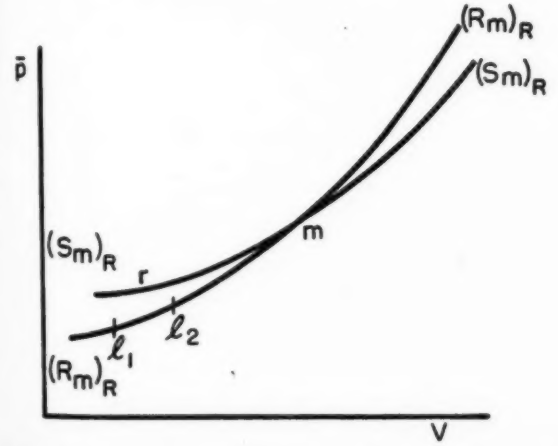


FIG. 7a. Velocity pressure diagram for rarefaction overtaking shock wave prior to interaction.

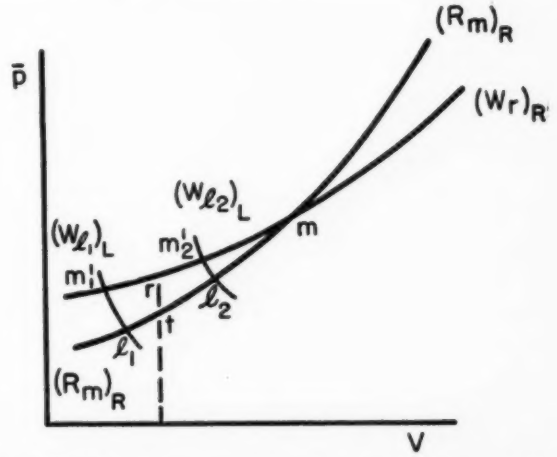


FIG. 7b. Velocity pressure diagram for rarefaction overtaking shock wave after interaction has taken place.

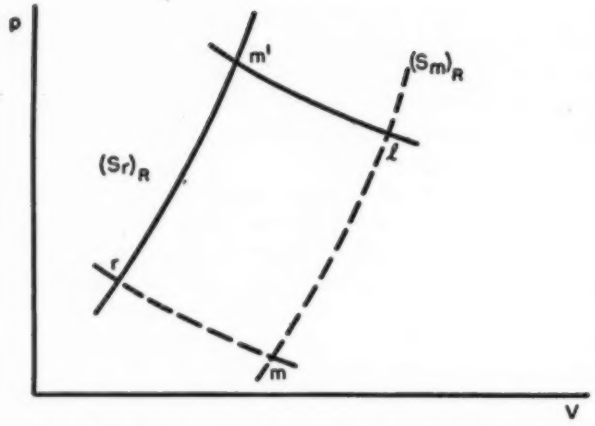


FIG. 8. Velocity pressure diagram for head-on collision of two shocks in arbitrary fluid.

may be shown to hold for any fluid. An example of such a property, which is common to all fluids with equations of state subject only to certain general restrictions, is the fact that two shock waves colliding head-on will always produce two shock waves after collision. We shall give here an outline of the proof of this particular result, which may be stated more precisely as follows.

Given a fluid with a caloric equation of state of the form $E=E(p, v)$ possessing the following thermodynamic characteristics, termed the Weyl conditions:[6]

1. Compression of the substance upon adiabatic application of external pressure, i.e., $(dv/dp)_{ad.}<0$;

2. Increased compression rate for adiabatic application of increased external pressure, i.e., $(d^2v/dp^2)>0$;

3. Unlimited pressure of the substance possible under continuous, adiabatic compression;

4. Single-valuedness of thermodynamic functions, i.e., the $p-v$ states form a convex region; and, in addition, satisfying the inequalities

5. $\partial E/\partial v\geqslant 0$, $\partial E/\partial p\geqslant 0$, $\partial^2 E/\partial v\partial p\geqslant 0$, $\partial^2 E/\partial v^2\leqslant 0$;

then the head-on collision of two shock waves will always produce two resulting shock waves, i.e.,

$$(S_1)_R(S_2)_L\rightarrow(S_3)_L(S_4)_R.$$

It is seen by direct differentiation, that an ideal gas, for which $E=(1/\gamma-1)pv$, satisfies the above inequalities in addition to the Weyl conditions and, hence, is a special case of such a fluid.

The proof is based on an argument similar to that used in the preceding section. It should be noted that v is a function of p (on any given shock line) and that the relation between these two variables is determined (once the caloric equation of state $E=E(p, v)$ is fixed) from the caloric shock equation

$$E(p, v)-E(p_0, v_0)=[(p+p_0)/2](v_0-v). \qquad (5)$$

[6] H. Weyl, *Shock Waves in Arbitrary Fluids* (Applied Mathematics Panel, NDRC AMP. Note No. 12, 1944).

We construct, therefore, a V, p (velocity-pressure) diagram (Fig. 8). Immediately before collision, the states in the three regions into which space is divided by the two approaching shock waves are related as follows:

$$p_m<p_l \quad \text{and} \quad p_m<p_r,$$
$$V_r<V_m<V_l.$$

To fix our ideas, assume that $p_r<p_l$ (a similar argument holds for $p_r>p_l$).

In order to show that two shock waves are always reflected as a result of the collision (i.e., that $p_{m'}>p_l>p_r$), it is sufficient to show that the shock curve $(S_r)_R$ through r will always lie to the left of the shock curve $(S_m)_R$ through m; i.e., that $(S_r)_R$ cannot intersect $(S_m)_R$.

Now the curve $(S_m)_R$ is given by the equation

$$V=V_m+\Phi_m(p);$$

whereas $(S_r)_R$ is given by

$$V=V_r+\Phi_r(p),$$

where from Eq. (4), $\Phi_0(p)=((p-p_0)(v_0-v))^{\frac{1}{2}}$. Since $V_r<V_m$, we must only show that for any given $p>p_r$

$$\Phi_r(p)<\Phi_m(p),$$

where $p_m<p_r$ and $v_m>v_r$. That $v_m>v_r$ is a consequence of a theorem[6] that v is always less than v_0 for $p>p_0$ on a shock line, and of the fact that r lies on a shock line through m.

It has also been shown that for a fluid with an equation of state satisfying the Weyl conditions,

$$L\equiv[\Phi_0(p)]^2=(p-p_0)(v_0-v)>0.$$

We shall accordingly deal with L and show that

$$L(p_m, v_m)>L(p_r, v_r)$$

for given p, and for $p_r>p_m$, $v_r<v_m$.

Consider L a function of p_0, v_0; now,

$$dL=(\partial L/\partial p_0)dp_0+(\partial L/\partial v_0)dv_0.$$

In order to insure a decrease in the value of L whenever there is an increase in p_0 and a decrease in v_0, it is required to show that

$$(\partial L/\partial p_0)<0 \quad \text{and} \quad (\partial L/\partial v_0)>0.$$

For an equation of state with the characteristics given above, we have

$$L=(p-p_0)[v_0-v(p_0, v_0, p)],$$

or

$$(\partial L/\partial p_0)=-(p-p_0)(\partial v/\partial p_0)-(v_0-v).$$

But from the caloric shock Eq. (5), we obtain

$$(\partial v/\partial p_0)=\frac{\frac{1}{2}(v-v_0)+(\partial E/\partial p)(p_0, v_0)}{(\partial E/\partial v)(p, v)+\frac{1}{2}(p+p_0)}.$$

Since $(\partial E/\partial v)$ and $(\partial E/\partial p)$ are positive on the basis of our assumptions and since $(v_0-v)>0$ by virtue of the

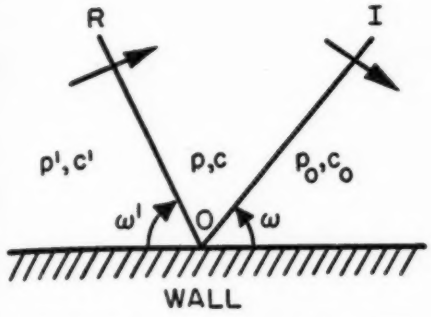


FIG. 9a. Regular reflection of shock wave from wall with coordinate system relative to wall.

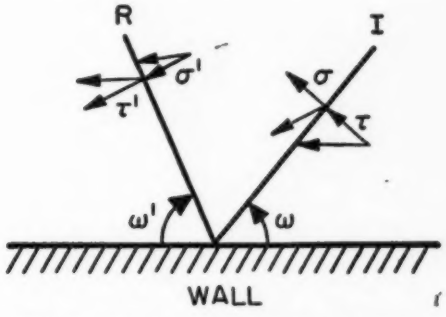


FIG. 9b. Regular reflection of shock wave from wall with coordinate system relative to line of contact.

previously stated theorem of Weyl,[6]

$$(\partial L/\partial p_0)=$$

$$-\left\{\frac{(p-p_0)[\frac{1}{2}(v_0-v)+(\partial E/\partial p)(p_0,v_0)]}{(\partial E/\partial v)(p,v)+\frac{1}{2}(p+p_0)}+(v_0-v)\right\}<0.$$

On the other hand,

$$(\partial L/\partial v_0)=(p-p_0)[1-(\partial v/\partial v_0)].$$

Since $p-p_0>0$, we need only prove that $(\partial v/\partial v_0)\leqslant 1$. From Eq. (5) we have

$$(\partial v/\partial v_0)=\frac{\frac{1}{2}(p+p_0)+(\partial E/\partial v)(p_0,v_0)}{\frac{1}{2}(p+p_0)+(\partial E/\partial v)(p,v)};$$

thus, $(\partial v/\partial v_0)\leqslant 1$, if $(\partial E/\partial v)(p,v)\geqslant(\partial E/\partial v)(p_0,v_0)$. Let $f\equiv\partial E/\partial v$. Then

$$df=(\partial f/\partial p)dp+(\partial f/\partial v)dv.$$

Since $p>p_0$ and $v<v_0$, $f(p,v)$ will always be greater than or equal to $f(p_0,v_0)$ provided that $\partial f/\partial p\geqslant 0$ and $\partial f/\partial v\leqslant 0$; i.e., provided that $\partial^2 E/\partial v\partial p\geqslant 0$ and $\partial^2 E/\partial v^2\leqslant 0$. Q. E. D.

V. SIMPLE THEORY OF REGULAR REFLECTION IN WATERLIKE SUBSTANCES

The oblique collision of two similar shock waves is mathematically equivalent to the oblique reflection of a single shock from an infinite plane rigid wall. Hence, consider a plane step-shock wave I (Fig. 9a) moving with constant velocity in a waterlike substance and incident at an angle ω on an infinite plane rigid wall. The reflection is said to be "regular"[1] if the line of contact O between the reflected plane shock R and the incident one is always on the wall. The phenomenon is evidently stationary for an observer moving with the line of contact (Fig. 9b), so that it is determined by the kinematical condition that the relative flow of the fluid is parallel to the wall both initially and finally. Now the only changes in material speed are normal to the shocks, *viz.*, v at the incident shock of strength ξ and v' at the reflected shock of strength ξ'. Since the total change of material speed normal to the wall must be zero, for normalization in the region between the two shocks,

$$v\cos\omega+v'\cos\omega'=0. \tag{14a}$$

Furthermore, the total change of material speed parallel to the wall must be equal to the difference between the initial parallel speed and the final one, i.e.,

$$v\sin\omega-v'\sin\omega'=\tau'/\sin\omega'-\tau/\sin\omega. \tag{14b}$$

Now τ and v are both functions of ξ: τ' and v' are both functions of ξ'. Hence, Eqs. (14a) and (14b) are sufficient to determine the strength ξ' of the reflected shock and its angle of reflection ω' in terms of the known strength ξ of the incident shock and its angle of incidence ω, if a solution exists. For computational purposes, it is convenient to combine these equations into a single quadratic equation as follows:

$$LX^2+MX+N=0, \tag{14}$$

where

$$X\equiv\cos^2\omega,$$
$$L\equiv v^2\{(\tau'^2-\sigma'^2)-(\tau^2-\sigma^2)\},$$
$$M\equiv v'^2\{(\tau^2-\sigma^2)+\sigma'^2\}-v^2\{(\tau'^2-\sigma'^2)+\sigma^2\},$$
$$N\equiv v'^2(\sigma^2-\sigma'^2).$$

The graphical representations of typical numerical solutions[2] of Eq. (14) are shown for $\gamma=7.15$ and given ξ [or p/p_0] in Fig. 10a for $\omega'(\omega)$, and in Fig. 10a′ for $\xi'(\omega)$]. Similar graphs for $\gamma=2.00$ are given in Figs. 10b and 10b′. As in the case of ideal gases,[1] there are, in general, two solutions ξ', ω' for given ξ, ω; one with large ξ' and large ω': the other with relatively small ξ' and small ω'. For certain "extreme" values (ω_e), however, the two solutions merge into one, while for still higher values of ω no solutions at all exist. The "extreme" values are determined by the condition that the derivative $\partial\xi'/\partial\omega$ or $\partial v'/\partial\omega$ becomes infinite (Fig. 10a). Subject to this condition, Eqs. (14a) and (14b) give for the "extreme" values $\omega(\xi')$

$$\tan^2\omega_e'=\left(\frac{\tau'}{v'}\frac{dv'}{d\sigma'}\right)_e=\frac{1}{\xi'^{1/\gamma}}\frac{1+f}{1-f},$$

where

$$f(\xi')\equiv(\xi'-1)/\gamma\xi'(\xi'^{1/\gamma}-1).$$

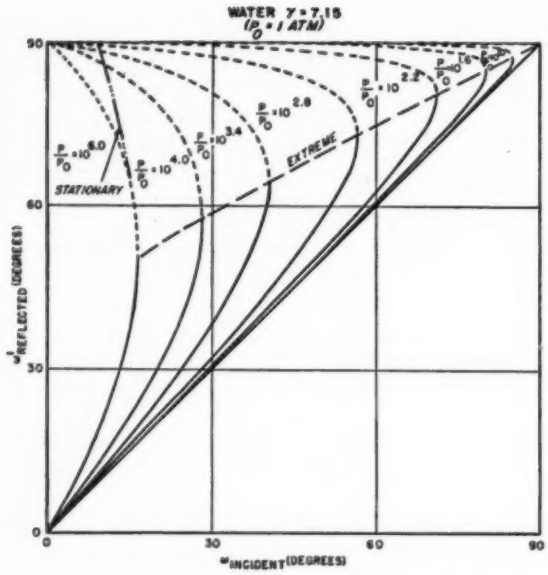


FIG. 10a. Regular reflection of shock waves, reflected angle *vs* angle of incidence, $\gamma = 7.15$.

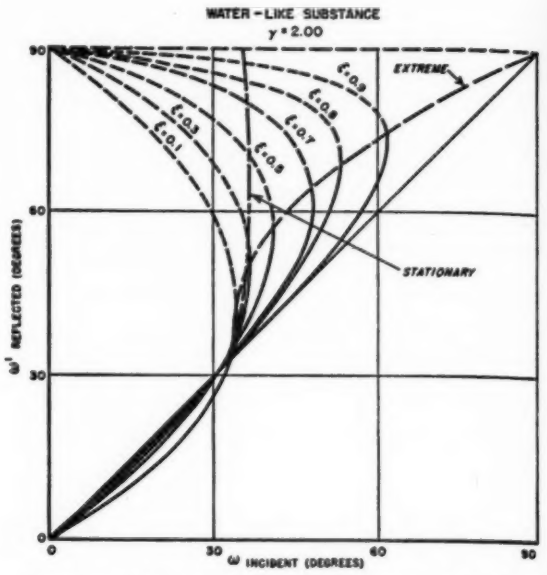


FIG. 10b. Regular reflection of shock waves, reflected angle *vs* angle of incidence, $\gamma = 2.00$.

Tables II and III list these "extreme" solutions ξ', ω' for $\gamma = 7.15$ and for $\gamma = 2.00$, 1.40 (included for comparison with air), respectively. Comparison shows that unless the shocks are very strong the limits of regular reflection for waterlike substances are similar to those for ideal gases having corresponding values of γ and obeying the Rankine-Hugoniot adiabatic.

The vanishing of the discriminant of Eq. (14) is sufficient to determine the value of ω for which the reflected

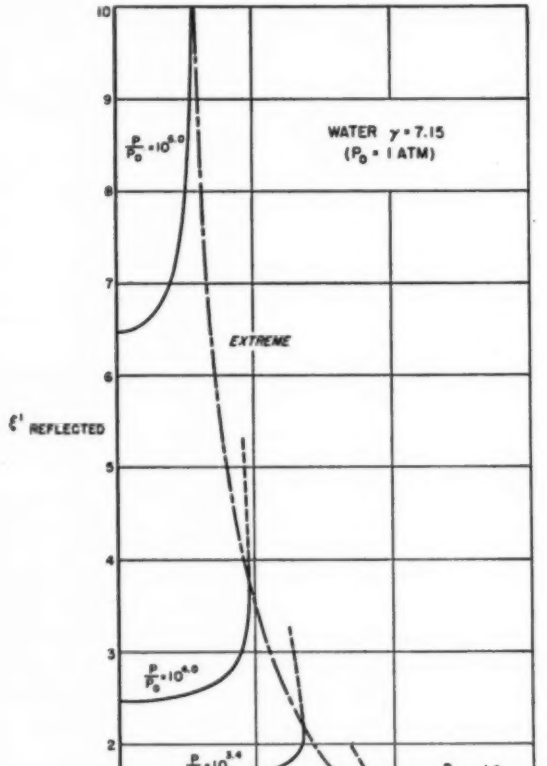


FIG. 10a'. Regular reflection of shock waves, reflected pressures *vs* angle of incidence, $\gamma = 7.15$.

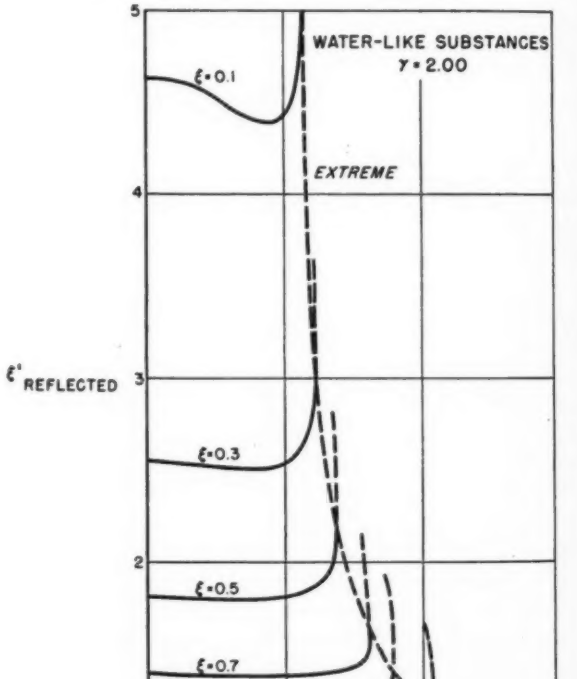


FIG. 10b'. Regular reflection of shock waves, reflected pressures *vs* angle of incidence, $\gamma = 2.00$.

pressure ξ' is a minimum for given ξ. Thus,

$$\cos^2\omega_m = -(N/L)^{\frac{1}{2}}.$$

The minimum solutions for $\gamma=2.00$ and 1.40 are given in Table IV; those that exist for $\gamma=7.15$ are not in the region of physical interest. The limiting value of ω for weak shocks ($\xi\rightarrow 1$) is the same as that for an ideal gas[1] having the same γ, namely,

$$\lim_{\xi\rightarrow 1} \sin^2\omega = 1-(\gamma+1)^{\frac{1}{2}}/2.$$

TABLE II. "Extreme" solutions.

		$\gamma=7.15$		
$\log_{10}(p/p_0)$*	ξ	ξ'	ω	ω'
0	1.0000	1.000	90°.00	90°.00
1.0	0.9970	1.006	85°.28	87°.64
1.6	0.9872	1.026	80°.29	85°.14
1.8	0.9797	1.041	77°.81	83°.89
2.0	0.9681	1.065	74°.80	82°.37
2.2	0.9501	1.102	71°.17	80°.52
2.4	0.9230	1.159	66°.90	78°.34
2.6	0.8831	1.247	62°.02	75°.82
2.8	0.8265	1.378	56°.66	73°.02
3.0	0.7502	1.569	51°.05	70°.05
3.2	0.6545	1.838	45°.48	67°.07
3.4	0.5444	2.197	40°.24	64°.24
3.6	0.4299	2.651	35°.57	61°.70
3.8	0.3224	3.190	31°.59	59°.53
4.0	0.2308	3.795	28°.31	57°.71
4.2	0.1592	4.442	25°.65	56°.23
4.4	0.1067	5.112	23°.53	55°.01
4.6	0.0701	5.793	21°.84	54°.00
4.8	0.0454	6.478	20°.49	53°.15
5.0	0.0291	7.167	19°.42	52°.41
6.0	0.0030	10.759	16°.43	49°.78

* $p_0=1$ atmos.

From Eq. (14a′) we note that the angle of incidence is equal to the angle of reflection provided

$$v' = -v. \tag{15a}$$

TABLE III. "Extreme" solutions.

	$\gamma=2.00$			$\gamma=1.40$		
ξ	ξ'	ω	ω'	ξ'	ω	ω'
0	∞	90°.00	0	∞	90°.00	0
0.1	5.291	34°.36	42°.31	5.455	41°.72	35°.08
0.2	3.874	35°.17	46°.87	3.845	40°.34	40°.96
0.3	3.118	36°.66	50°.48	3.051	40°.42	45°.52
0.4	2.605	38°.59	53°.80	2.537	41°.27	49°.63
0.5	2.218	40°.98	57°.10	2.160	42°.76	53°.62
0.6	1.905	43°.98	60°.55	1.862	44°.95	57°.69
0.7	1.641	47°.87	64°.35	1.612	48°.10	62°.08
0.8	1.410	53°.27	68°.83	1.395	52°.83	67°.12
0.9	1.200	61°.83	74°.76	1.195	60°.97	73°.63
1.0	1.000	90°.00	90°.00	1.000	90°.00	90°.00

Now this relation always holds for the one-dimensional case of head-on reflection ($\omega_h=\omega_h'=0$). The head-on values of ξ_h' for $\gamma=7.15$ are given in Table V; and for $\gamma=2.00$ and 1.40, in Table VI. Furthermore, substituting condition (15) in Eq. (14b), we find that for the same ξ_h' the angle of incidence (ω_h) equals the angle of reflection also if

$$\sin^2\omega_h = (\tau'-\tau)/2v. \tag{15b}$$

TABLE IV. "Minimum" solutions.

	$\gamma=2.00$			$\gamma=1.40$		
ξ	ξ'	ω	ω'	ξ'	ω	ω'
0.1	4.393	26°.92	20°.95	4.946	38°.73	22°.64
0.2	3.139	24°.90	20°.87	3.371	35°.35	24°.00
0.3	2.504	23°.83	20°.86	2.622	33°.44	24°.87
0.4	2.094	23°.14	20°.88	2.156	32°.16	25°.55
0.5	1.796	22°.64	20°.94	1.830	31°.17	26°.12
0.6	1.568	22°.27	21°.02	1.585	30°.39	26°.63
0.7	1.385	22°.00	21°.11	1.393	29°.76	27°.10
0.8	1.234	21°.79	21°.22	1.237	29°.22	27°.54
0.9	1.108	21°.58	21°.32	1.108	28°.78	27°.96
1.0	1.000	21°.47	21°.47	1.000	28°.34	28°.34

Inasmuch as v is intrinsically negative for $0\leqslant\xi\leqslant 1$, no such angle exists for $\tau'>\tau$; hence, the limiting condition is given by $\tau'=\tau$, for which $\omega_h=0$.

It is seen that the magnitude of ω_h for waterlike substances varies not only with the value of γ, as is the case of ideal gases, but also with the strength ξ of the incident shock. To understand this difference we refer to Fig. 11, where $v'(\xi')$ is plotted against $\tau'(\xi')$ and $-v(\xi)$ against $\tau(\xi)$ for $\gamma=1.40$, 2.00, and 7.15. Note that for $\gamma=3$, the curves form a vertical cusp in the neighborhood of weak shocks. For $\gamma\leqslant 3$, the curves intersect at a point $\tau'=\tau$ only for weak shocks, i.e., $\tau'\sim\tau\sim 1$. Let $\xi\rightarrow 1$. Then

$$-v\sim 1-(1-\xi)/\gamma,$$

TABLE V. "Head-on" solutions.

	$\gamma=7.15$
$\log_{10}(p/p_0)$*	ξ_h'
0.0	1.000
1.0	1.003
1.6	1.013
1.8	1.020
2.0	1.032
2.2	1.051
2.4	1.080
2.6	1.125
2.8	1.192
3.0	1.291
3.2	1.431
3.4	1.620
3.6	1.860
4.0	2.472
5.0	4.344
6.0	6.449

* $p_0=1$ atmos.

TABLE VI. "Head-on" solutions.

ξ	$\gamma=2.00$ ξ_h'	$\gamma=1.40$ ξ_h'
0.0	∞	∞
0.1	6.171	4.634
0.2	3.806	3.229
0.3	2.823	2.548
0.4	2.258	2.116
0.5	1.882	1.808
0.6	1.610	1.574
0.7	1.404	1.388
0.8	1.241	1.235
0.9	1.109	1.108
1.0	1.000	1.000

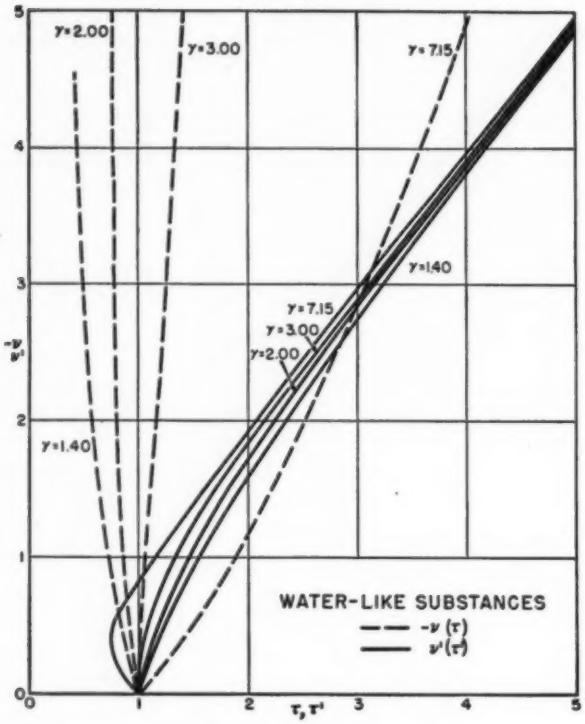


FIG. 11. Plot of $\nu'(\xi')$ *vs* $\tau'(\xi')$.

and

$$\tau \sim 1-[(\gamma-3)/4\gamma](1-\xi);$$

hence, $d(-v)/d\tau \sim 4/\gamma-3$, which is independent of ξ (weak) and which becomes infinite for $\gamma=3$. Furthermore, in this γ range $\tau'<\tau$ so that the angle ω_A exists for all strengths ξ. On the other hand, for $\gamma>3$ (e.g., $\gamma=7.15$) the curves start out in a reversed position so that τ' is initially greater than τ; in other words, there is no real angle ω_A. At some lower value ξ_A, however, the curves cross over so that then $\tau'<\tau$ and ω_A exists. For example, for $\gamma=7.15$, $\xi_A=8.48(10)^{14}$ is the weakest shock for which $\omega=\omega'$. This value of $p_{0A}/p_A=3.538(10)^{16}$ lies outside the domain of physical interest.

For angles of incidence greater than ω_A or 0, if ω_A does not exist, the reflected pressure always exceeds the head-on value. On the other hand, the extreme value of ξ_e' may coincide with the head-on value for a particular strength ξ_A of the incident shock. In that case head-on reflection gives higher values of ξ' than any oblique reflection for all stronger incident shocks ($\xi<\xi_{eA}$). Table VII gives ξ_{eA} for $\gamma=7.15$, 2.00, 1.40.

For a given strength of the incident shock I, there is always one angle of incidence ω_e for which the pressure and flow direction of the material behind the reflected shock R are identical with those corresponding to the simple theory of a three-shock intersection (Fig. 12). This "quasi-stationary" flow corresponds to the regular reflection solution with low-valued ξ_e' for weak incident shocks, but with the high-valued ξ_e' for strong incident shocks. The critical strength ξ_{ee} that separates these two classes of quasi-stationary solutions is that particular stationary one which is identical with an "extreme" solution. J. von Neumann[7] has determined the "extreme-stationary" equality by means of an asymptotic expansion for relatively strong shocks ($\xi\sim 0$) in which terms like ξ are neglected, but not terms such as $\xi^{1/\gamma}$. It turns out that ξ_{ee}' must satisfy the following equations:

$$\frac{(F\xi'^{(\gamma+1)/\gamma}-(F\xi'^{(1/\gamma)}-1))^2}{F\xi'^{(1/\gamma)}-1} = \frac{4\xi'^{(\gamma+2)/\gamma}(\xi'-1)}{(\xi'^{(1/\gamma)}-1)(1+f)\{\xi'^{(1/\gamma)}(1-f)+(1+f)\}},$$

where

$$f(\xi')\equiv\frac{\xi'-1}{\gamma\xi'(\xi'^{(1/\gamma)}-1)},$$

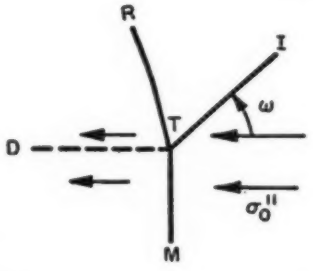


FIG. 12. Triple-shock configuration.

and

$$F(\xi')\equiv\frac{(1-f)(\xi'^{(1/\gamma)}-1)\xi'}{\xi'^{(1/\gamma)}(1-f)+(1+f)}+\frac{\xi'^{(\gamma+1)/\gamma}-1}{\xi'^{(1/\gamma)}(\xi'-1)}=\xi^{-1/\gamma}.$$

Thus, for $\gamma=7.15$ one obtains the asymptotic value (Table VIII) $\xi_{ee}=0.0001435$, $\xi_{ee}'=16.395$, $\omega_{ee}=15°.24$, $\omega_{ee}'=47°.46$. The existence of such a contact for all γ in the case of waterlike substances is still another respect in which these behave differently from the analogous ideal gases,[1] for which no such phenomena occurs when $\gamma>3.5931$.

Experimental verification[2] of the interaction of shock waves from explosions in water was initiated in 1943 by P. Libessart and continued independently by D. C. Campbell and by D. P. MacDougall, H. Messerly, and E. M. Boggs. An exploratory quantitative investigation was carried out in 1944 by the group at the underwater Explosives Research Laboratory, Woods Hole, Massachusetts. These limited experimental observations show qualitative agreement with the simple theory of regular reflection and the phenomenon of Mach reflection where regular reflection does not exist. A comprehensive survey is still needed.

TABLE VII. "Extreme-head-on" equality.

γ	ξ_{eA}	ξ_{eA}'	ω_{eA}	ω_{eA}'
7.15	$6.243\times(10)^{-25}$	3563.	31°.08	31°.08
2.00	0.02345	9.083	35°.80	35°.80
1.40	0.1889	3.965	40°.39	40°.39

[7] J. von Neumann, Private communication to R. J. Seeger, 24 November 1943.

VI. SIMPLE THEORY OF THREE-SHOCK INTERSECTIONS IN WATERLIKE SUBSTANCES

Three plane shocks I, R, M intersect along a line 0 (normal to the plane of the paper in Fig. 13. The simple theory[1] of their intersection requires that each region between adjacent shock have uniform pressure. Although subsequent experimental investigations in air have show that in the case of a developing Mach effect this condition is valid only for the regions between I and R and between I and M, the whole phenomenon is so little understood at present that a survey of three-shock configurations in waterlike substances still has heuristic value. Consideration of the fluid flow relative to the moving intersection line 0 indicates that one of these regions must have a plane D along which there is a discontinuity D in tangential material speed. This "slipstream" for waterlike substances, of course, is not a density discontinuity (in contrast to the case of an ideal gas).

Numerical results for $\omega(\beta)$ are shown graphically in Fig. 14 for $\gamma=7.15$. In general, the various results for strong incident shocks ($\xi=0.75$ and $\xi=0.10$) agree approximately with those for weak shocks ($\xi\rightarrow 1$).

Of particular interest is the case where the material flow across shock M is normal to it. Let σ'' refer to the flow in the region behind M, normalized with respect to the region ahead of M. It can be shown that

$$\sigma''(\xi/\xi')=\sigma'(1/\xi)\cos\omega.$$

This relation must hold in addition to the general Eqs. (14a) and (14b). Hence, for a given value of ξ a particular configuration (ω_s, ξ_s', ω_s') is determined. Tables IX and X give these so-called "stationary" values for $\gamma=7.15$ and for $\gamma=2.00$, 1.40, respectively. In the limiting case of weak shocks we obtain

$$\lim_{\xi\rightarrow 1}\sin\omega_{s1}=(\gamma/2(2\gamma-1))^{\frac{1}{2}},$$

and

$$\lim_{\xi\rightarrow 1}\xi_{s1}'=2\gamma.$$

For the limit of weak incident shocks, i.e., $\xi\rightarrow 1$, the various conditions result in the following cubic equation, determining the angle β_1:

$$\{4\sigma'\cos\beta_1-[(\gamma+1)r\sigma'-2(\gamma-1)\sigma'^2]\}\times\{[\sigma'^2-\tau'^2]\cos^2\beta_1-2\sigma'\cos\beta_1+(\tau'^2+1)\}-[(m\sigma'^2+n)\cos^2\beta_1-2m\sigma'\cos\beta_1+(m-n)]=0,$$

where

$$m\equiv(\gamma+1)s\sigma'+2(\gamma-1)\sigma'\tau',$$
$$n\equiv\sigma'\tau'\{(\gamma+1)r\sigma'-2(\gamma-1)\sigma'^2-4\},$$
$$r\equiv[4\gamma/(\gamma+1)]\xi'(d\sigma'/d\xi'),$$

TABLE VIII. "Extreme-stationary" equality.

γ	$\xi_{s\infty}$	$\xi_{s\infty}'$	$\omega_{s\infty}$	$\omega_{s\infty}'$
7.15	0.0001435	16.3948	15°.24	47°.46
2.00	0.2514	3.442	35°.87	48°.78
1.40	0.4137	2.479	41°.44	50°.18

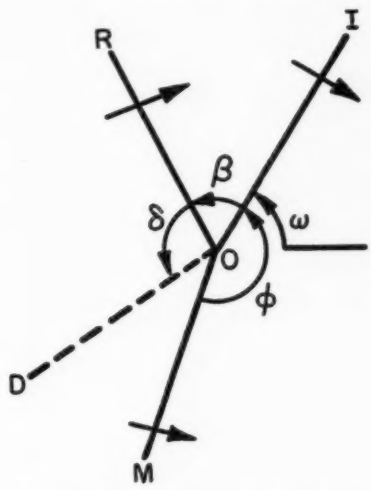


FIG. 13. Triple-shock configuration with system of coordinates fixed at triple point.

and

$$s\equiv-[4/(\gamma+1)]\xi'(d\tau'/d\xi').$$

The other angles are then given by

$$\delta_1=\pi-\cot^{-1}[(\sigma'\cos\beta_1-1)/\tau'\sin\beta_1],$$

and

$$\phi_1=\pi-\beta_1,$$

and

$$\omega_1=\cot^{-1}[(\sigma'-\cos\beta_1)/\sin\beta_1].$$

For the case of weak reflected shocks, $\xi'\rightarrow 1$, the following quadratic equation, analogous to that for ideal gases, may be derived in terms of the cosine of the limiting angle β_1':[1]

$$\sigma\{\sigma^2(\tau-\sigma)-(3\sigma+\tau)\}\cos^2\beta_1'+4\sigma(\sigma^2+1)\cos\beta_1'+\{\sigma\tau(1-\sigma^2)-(1+3\sigma^2)\}=0.$$

As is the case for ideal gases ϕ_1' approaches π as a limit, and the limiting angles ω_1', δ_1' are given by

$$\tan\omega_1'=\tau/(\csc\beta_1'-\tau\cot\beta_1'),$$

and

$$\cot\delta_1'=\csc\beta_1'-\cot\beta_1'.$$

An interesting distinction between ideal gases and waterlike substances becomes apparent here when one consider a value of ξ slightly less than one. Let this correspond to a value of $\sigma=1-\epsilon$, where ϵ is a small

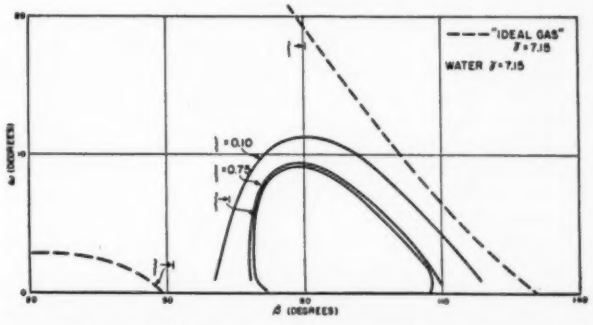


FIG. 14. Triple-shock configurations, plot of ω vs β.

TABLE IX. "Stationary" solutions.

$\log_{10}(p/p_0)$*	$\gamma=7.15$ ξ'
0.0	142.0
1.0	141.8
1.6	141.1
1.8	140.6
2.0	139.8
2.2	138.5
2.4	136.6
2.6	133.7
2.8	129.5
3.0	123.7
3.2	116.0
3.4	106.4
3.6	95.57
3.8	84.10
4.0	72.87
4.2	62.53
4.4	53.46
4.6	45.77
4.8	49.41
5.0	34.23
6.0	20.35

* $p_0=1$ atmos.

TABLE X. "Stationary" solutions.

ξ	$\gamma=2.00$ ξ'	$\gamma=1.40$ γ'
0.0	∞	∞
0.1	4.441	5.444
0.2	3.583	3.373
0.3	3.379	2.729
0.4	3.361	2.496
0.5	3.418	2.422
0.6	3.509	2.418
0.7	3.619	2.450
0.8	3.740	2.502
0.9	3.868	2.567
1.0	4.000	2.639

positive quantity. The discriminant of the above quadratic equation is then given, in terms of a power series in ϵ, by the expression

$$16\epsilon^4\{[(\gamma-3)/(\gamma+1)-1]+(\text{terms in }\epsilon)\},$$

which, for $\gamma>1$, is negative for a sufficiently small value of ϵ. Hence, for sufficiently weak incident shock there is always a region of values for which no real three-shock solutions (with correspondingly weak reflected shocks) can be found in the case of waterlike substances.

For ideal gases, on the other hand, the discriminant is given by the analogous expression

$$64\epsilon^4[(\gamma^2-1)-2(\gamma-1)^2\epsilon],$$

which, for $\gamma>1$, is positive for a sufficiently small value of ϵ.

In this instance, for sufficiently weak incident shocks, there is always a region of values for which real three-shock solutions (with correspondingly weak reflected shock) must exist. In fact, as $\xi\to1$, $\beta_1'\to0$, $\phi_1'\to\pi$, $\tan\omega_1'\to\{1\pm2[(\gamma-1)/(\gamma+1)]^{\frac{1}{2}}\}^{\frac{1}{2}}$, and $\delta_1'\to\pi-\omega_1'$.

These results are quite surprising. Even though the incident shock is very weak ($\xi\to1$) the minimum resultant pressure $\xi_1'_{\min}$, may be many times greater, e.g., $\xi_1'_{\min}=2.0234$ for $\gamma=1.40$, $\xi_1'_{\min}=3.3175$ for $\gamma=2.00$, and $\xi_1'_{\min}=138.67$ for $\gamma=7.15$. In other words, for weaker resultant pressures ξ_1' no three-shock configurations exist at all; they do exist, however, for an ideal gas with the same value of γ. Numerical results for $\xi=0.75$, 0.10 ($\gamma=7.15$) show that also for these values of ξ, there exist minimum values for the resultant pressures $\xi'_{\min}$ (e.g., $\xi'_{\min}=120.2$ for $\xi=0.75$ and $\xi'_{\min}=48.1$ for $\xi=0.10$), in contrast to the three-shock configurations with weak reflected shock for the analogous ideal gas.

It should be remarked that the more rigorous analysis of V. Bargmann,[8] valid only for weak shocks in the case of ideal gases, holds for arbitrary shock strength in the case of waterlike substances. Further results on this question will be reported elsewhere.

The authors would like to express their appreciation for the invaluable assistance rendered by Arnold George Rawling in preparing and checking the manuscript.

[8] V. Bargmann, *On Nearly Glancing Reflection of Shocks* (Applied Mathematics Panel NDRC AMP Report 108.2R, 1945).

Summer Program at Oak Ridge Institute of Nuclear Studies

Three additional basic courses in radioisotope techniques of four weeks duration and a three-week autoradiography course will be offered this summer by the Special Training Division of the Oak Ridge Institute of Nuclear Studies.

The basic courses will be the twenty-first, twenty-second, and twenty-third to be given by the Institute since the program was initiated in June, 1948. They are scheduled to begin on June 11, July 9, and August 15. The courses combine lectures, demonstrations, and laboratory work. Sufficient space is available to permit individual laboratory work. Thirty-two participants will be accepted for each course.

The autoradiography course, which will begin on July 2, will be the first of its kind to be offered by the Institute. It is intended for personnel who will direct medical or biological research utilizing the autoradiographic process.

George A. Boyd, formerly of the University of Rochester, who is well known for his work in autoradiography, will direct the course. Subjects to be covered include the theory of the photographic process, reaction of ionizing particles with photographic emulsions, and the interpretation of results, and techniques of making gross and microscopic autoradiograms. Twenty participants will be accepted for the course.

Lecturers and laboratory leaders of the course will include:

Julian Webb and John Spence, Eastman Kodak Company.
Robert Dudley, Massachusetts Institute of Technology.
Margaret Holt, New England Deaconess Hospital.
C. P. Leblond and Rita Bogorach, McGill University.
L. F. Belanger, University of Ottawa.
S. R. Pelc, Hammersmith Hospital, London.
Agnes Williams, University of New Mexico.
Harvey Blank, University of Pennsylvania.

Registration is $25 for each course. Additional information and application forms are available from Ralph T. Overman, Chairman, Special Training Division, Oak Ridge Institute of Nuclear Studies, P.O. Box 117, Oak Ridge, Tennessee.

Scattering Phenomena in Electron Microscope Image Formation

C. E. HALL
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 8, 1950)

Equations are derived for the relative intensities in electron microscope images formed without a limiting aperture in the objective lens. Three components of intensity at the image plane can be distinguished: I_T, a transmitted intensity of very small relative aperture; I_K, an intensity scattered within the effective aperture of the objective lens; and I_B, a background intensity which does not contribute to the recognizable image. The equations derived involve two scattering cross sections, both of which are readily measurable under normal operating conditions. The form of the equations is verified experimentally; and cross sections are measured for Be, SiO, Cr, Ge, Pd, Pt, and U. The effect of altering the beam potential is discussed in the light of the results and the minimum detectable increment in specimen thickness due to scattering is estimated.

I. THEORY

THE formulas to be derived are based on semi-quantitative considerations in an attempt to account for image intensities in terms of empirical scattering constants, readily measurable with the electron microscope itself under ordinary operating conditions. In recognition of the fact that highest resolution can be obtained without incorporating a limiting aperture in the objective lens, it is assumed that the physical aperture of the objective is appreciably greater than the effective aperture of the lens $(d/Cf)^{\frac{1}{4}}$, where d is the distance between resolvable centers in the specimen and Cf is the product of the dimensionless spherical aberration constant and the focal length. Objective diaphragms are considered to be of secondary importance in image formation and their purpose may be judged in the light of experimental results. Since it is the almost universal practice to record electron micrographs with a relative aperture of illumination of the order of 10^{-4} radian, the incident intensity may be considered as a plane wave. Two diffracting centers in the object are resolvable if the first-order maxima are contained within the angle α, so that $\alpha > \lambda/d$. The smallest d is then $d_0 \simeq (Cf)^{\frac{1}{4}}\lambda^{\frac{3}{4}}$, which is essentially the same resolving power as would be obtained with an aperture stop of optimum diameter.

Since the spherical aberration error increases as α^3, electrons scattered through angles appreciably greater than the effective aperture of the lens reach the image plane so far removed from their proper position in the image that they appear only as a diffuse background. Similarly, electrons having suffered large energy losses will, if they reach the image plane, also contribute to the background intensity. Intensities in the recognizable image consist of two parts: (a) a scattered intensity, I_K, contained within the effective aperture of the objective, distinguishable from the background owing to the sharp increase in intensity at small angles, and (b) a transmitted intensity I_T, of aperture approaching that of the illuminating beam. The intensity I_T produces a high resolution image characterized by the familiar fresnel fringes. Superimposed is an image of inferior quality produced by I_K and similar to those visible in dark-field. It is not strictly accurate to draw a sharp distinction between the three components, ignoring the continuous angular distribution of intensity; but in making these assumptions we arrive at relatively simple formulas describing the main characteristics of electron microscope images within reasonable limits of error. Let σ_T be the atomic cross section in cm² for scattering of all kinds out of the transmitted beam and define the corresponding total cross section per gram as $S_T = N_A \cdot \sigma_T/A$ where N_A is Avogadro's number and A is the atomic weight. Let w be the mass thickness, or mass per cm² of specimen. Then,

$$dI_T = -I_T S_T dw, \tag{1}$$

and

$$I_T = I_0 e^{-S_T w}. \tag{2}$$

Of the intensity dI_T lost to the transmitted wave in the layer dw, we assume a fraction k is scattered within the effective aperture and so will contribute to I_K. The rest of the scattered intensity $(1-k)dI_T$ is permanently lost to the useful imaging beam. In a layer dw, I_K will thus be increased by an amount kdI_T and also decreased by an amount $I_K(1-k)S_T dw$. With due regard to the fact that dI_T is a negative number, we have

$$dI_K = -kdI_T - I_K(1-k)S_T dw. \tag{3}$$

When dI_T is replaced through relations (1) and (2), Eq. (3) is readily integrated to give Eq. (6) below; but it is a little easier to follow another course. Define I_c, the total coherent intensity at the image plane, such that $I_c = I_T + I_K$. Using this relation and substituting for I_T in Eq. (3), we have

$$dI_c = -I_c(1-k)S_T dw, \tag{4}$$

which, with $(1-k)S_T = S_c$, becomes

$$I_c = I_0 e^{-S_c w}. \tag{5}$$

Then it follows that

$$I_K = I_0 e^{-(1-k)S_T w} - I_0 e^{-S_T w}. \tag{6}$$

The atomic cross section for incoherent scattering is

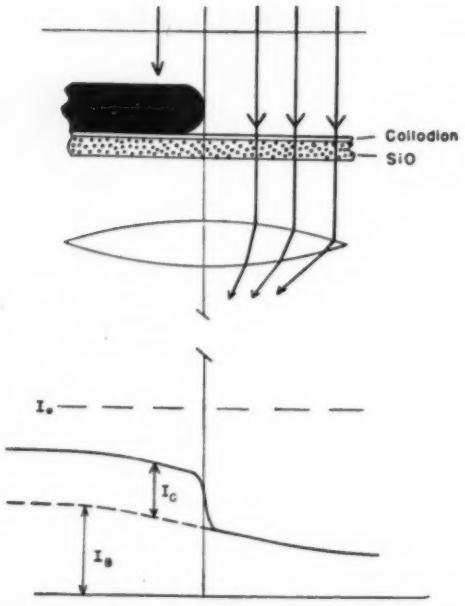


FIG. 1. Schematic diagram showing method of measuring relative magnitude of I_c.

$\sigma_e = S_e A/N_A$. Note that the term "coherent" as used here is relative to the apparatus, since wide-angle diffractions are excluded from I_c. Equations (2), (5), and (6), involving two constants, specify the magnitude and character of the image intensity (superimposed on I_B); and it is desirable to test the equations experimentally.

The various numerical constants of the magnetic objective lens used in the experimental part are listed in Table I. No physical aperture was employed. The spherical aberration length, Cf, had been measured from the displacement of diffraction images,[1] and asymmetries had been compensated after the method of Hillier and Ramberg.[2] The microscope is an RCA Type B.

II. TEST OF THE EXPONENTIAL LAW FOR I_c

In order to test Eq. (5) for I_c, it was necessary to produce a series of films of known relative thickness. SiO* was chosen as the material for the purpose because it is noncrystalline and had been used extensively on previous occasions with reproducible results for supporting films and replicas. Since geometrical calculations of mass thickness of films produced by vacuum evaporation are unreliable and since the absolute value of w is not required, all depositions were made at the same filament-to-specimen distance, and from the same filament on the assumption that under similar conditions of geometry, vacuum, etc., the amount of SiO deposited is proportional to the amount evaporated. Weighed amounts of SiO were deposited from a distance of about 11 cm on Lektromesh grids carrying collodion films, and grids were interchanged between evaporations to obtain varying combinations of evaporated amounts. An absolute calibration of mass per cm² was obtained from the increase in weight produced by deposition of some of the thicker films on glass cover slips situated beside the grids. The conversion factor is, 0.8×10^{-6} g per cm² per mg evaporated.

TABLE I. Objective lens (compensated).

$\lambda = 0.047$ A	$\alpha_0 = 5.\times10^{-3}$ radian
$V = 65$ kv	Bore, object side = 4.9 mm
$f = 3.5$ mm	Bore, image side = 2.3 mm
$Cf = 0.56$ cm	Gap = 2.0 mm
$d_0 = 10$ A	

The ratio, I_c/I_0, for various film thicknesses was obtained from a record of the intensity near the edge of a grid as illustrated in Fig. 1. On the bright side of the edge, the intensity is I_c+I_B, and inside the edge it is I_B only. The drop in intensity at the edge is therefore a measure of I_c as indicated in the diagram. I_0 must also be obtained through an open area, since it is the ratio I_c/I_0 which is required.

The experimental procedures were as follows: Films were broken away with a sharp needle from some of the squares in a grid and the specimen was placed in the microscope with the film side down as shown in Fig. 1. At a magnification of 20,000×, the microscope was focused on the plane of the film at the edge of a square. The aperture of illumination was then reduced with the condenser lens control and a 2×3 inch frame exposed

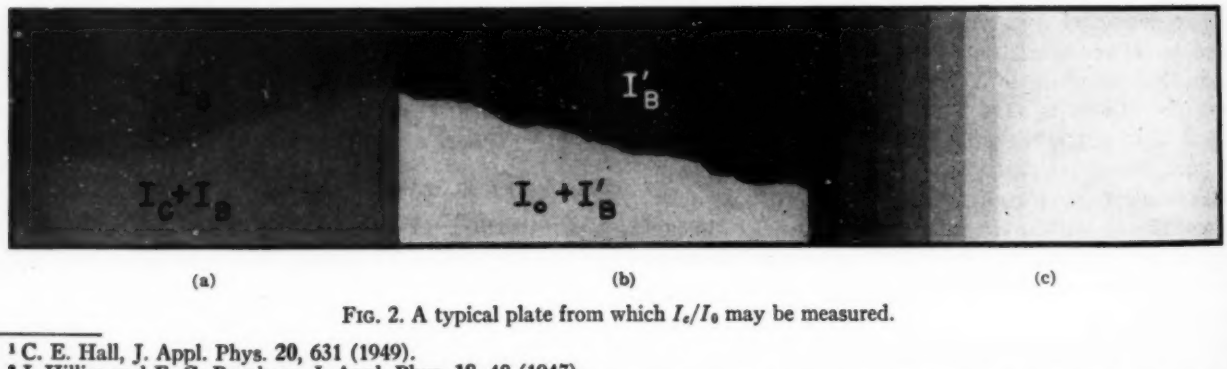


FIG. 2. A typical plate from which I_c/I_0 may be measured.

[1] C. E. Hall, J. Appl. Phys. 20, 631 (1949).
[2] J. Hillier and E. G. Ramberg, J. Appl. Phys. 18, 48 (1947).
* The author is indebted to the Kemet Division, Union Carbide and Carbon Corporation, for a typical analysis of this material: SiO—greater than 95 percent, Mn—0.01 percent, Na—0.10 percent, C—trace, Si plus SiO_2 balance.

for 20 sec as shown in Fig. 2(a). Without alteration of incident-intensity, the edge of an adjacent square from which the film had been removed was brought into the field and the second frame was exposed for from 10 to 20 sec as in Fig. 2(b). With the central region of the empty square brought into the field, and the plate shifted to the third frame, the illumination was reduced to expand the time scale. The shutter was then opened and the plate shifted at intervals of 10 sec to obtain steps from which a calibration curve of photographic density as a function of relative exposure could be made as shown in Fig. 2(c). Note that I_B in the first frame is not the same as I_B in the second frame, because the latter is distant from scattering areas. Plates were

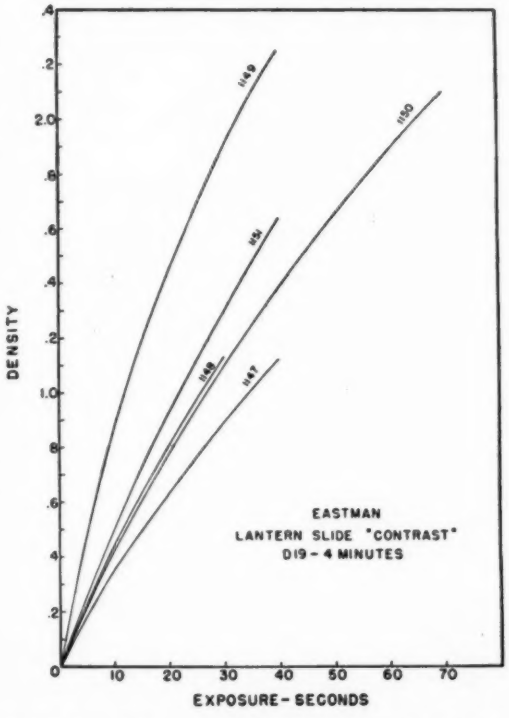


FIG. 3. A set of typical calibration curves obtained from stepped exposures as shown in Fig. 2(c).

Eastman Lantern Slide Contrast, developed in D19 for 4 minutes at a temperature between 68 and 70°F.

A microphotometer† tracing with a slit-width corresponding to about 100A was plotted across the edge in Frame 1 from visual readings of a density $[\log_{10}(1/T)]$ scale. Densities were read at a number of places on either side of the edge in Frame 2 and the average obtained. Similarly, the densities of calibration steps were measured and plotted for each plate as a function of the exposure time, as shown in Fig. 3, from which curves the required intensities are obtained for the densities in Frames 1 and 2. Plotted points are not shown in Fig. 3, because they fall on the lines as drawn. A typical tracing from Frame 1, replotted to show in-

†Sinclair Smith Recording Microphotometer, Fred C. Henson Company, Pasadena, California.

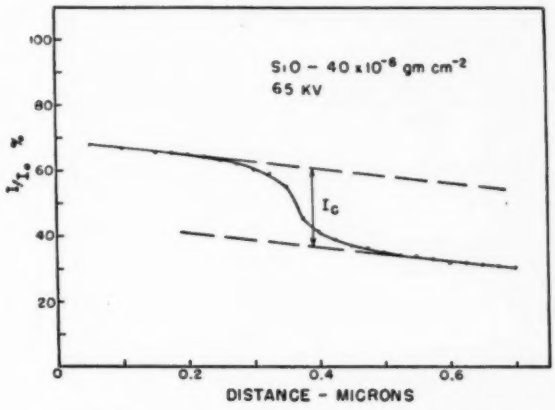


FIG. 4. Microphotometer tracing across an edge as recorded in Fig. 2(a).

tensity as percent of I_0, is reproduced in Fig. 4. On either side of the edge, the intensity slopes because of a gradual decrease in background into the shadow of the edge. Also, the edge is slightly out of focus because it is not at the same level as the film. I_e was measured as indicated between the projections of the curve on either side of the edge and at a point about half-way across the transition region. A series of edge traces from films of increasing thickness are shown in Fig. 5, where the numbers above the traces are proportional to the thickness and correspond to the weight of evaporated SiO. For films much thicker than the thickest shown, the intensity is practically the same on either side of the edge.

Results from the I_e experiments are shown in Fig. 6, where $\log_e(I_0/I_e)$ is plotted against the relative film thickness. The results are satisfactorily represented by a linear relationship, verifying the form of Eq. (5) within experimental errors, of which the largest source is probably uncertainty regarding the relative thickness of the films. The fitted straight line does not go through zero because of the underlying collodion film which is judged to have a scattering power equivalent to about 1 to 2 mg of evaporated SiO. On the right side of the graph in Fig. 6, a scale is drawn showing I_e/I_0 directly, demonstrating that the exponential law holds sur-

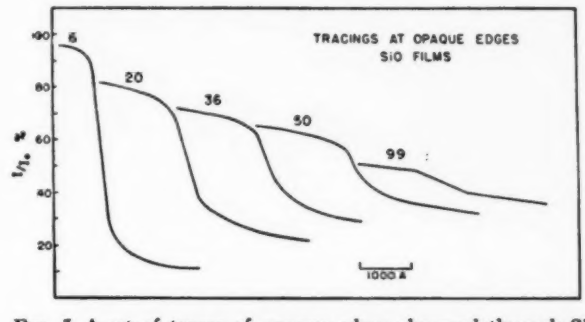


FIG. 5. A set of traces of opaque edges observed through SiO films of mass thickness proportional to the numbers over the traces. Multiply by 0.8×10^{-6} g cm^{-2} to convert the relative numbers to absolute values.

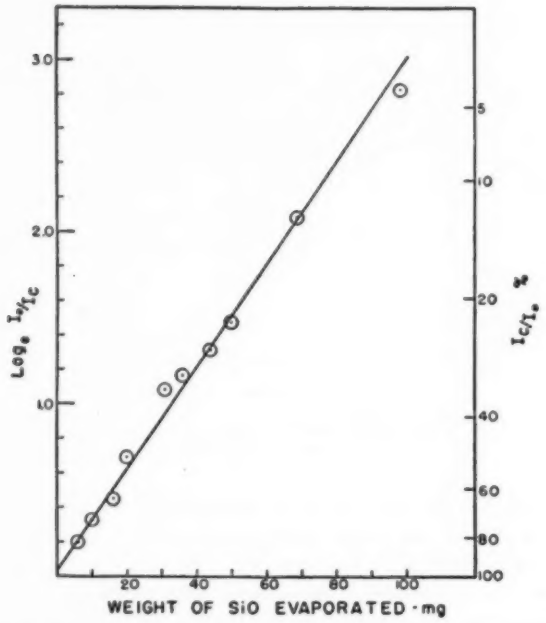


FIG. 6. Experimental test of the exponential law for I_c, Eq. (5). The abscissas are relative thickness and are convertible to absolute values when multiplied by 0.8×10^{-6} g cm^{-2}.

prisingly to values of I_c/I_0 of about 5 percent. Resolution is, of course, poor through such thick films; but the low visibility of the edge is due mainly to the high background intensity. When the abscissas in Fig. 6 are converted to absolute values, the slope of the line is $S_c=3.7\times10^4$ cm^2 g^{-1}.

Another plot of the data is shown in Fig. 7, where $I_s=I_0-I_c$ is that part of the scattered intensity which does not contribute to the useful image and I_B is that part of I_s reaching the image plane as background. The vertical distance between curves represents the part of I_s lost against the walls and apertures of the instrument. Note that with the wide aperture used, a very high proportion of the intensity, I_s, reaches the image plane as background up to film thicknesses of about

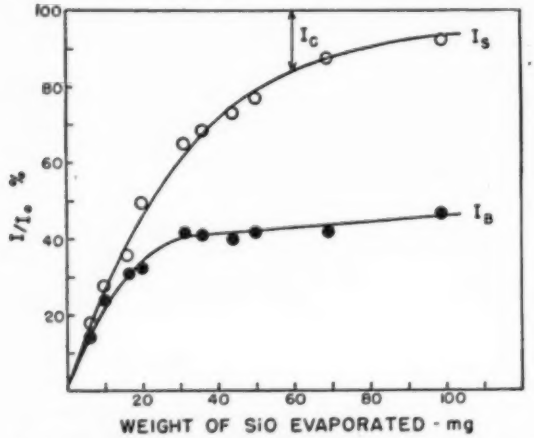


FIG. 7. Graphs showing the relative magnitudes of I_c, I_s, and I_B with increasing thickness of SiO films.

FIG. 8. Geometry for the experimental determination of I_T/I_0. Region D.F. appears in dark-field and B.F. in bright-field. The spacer used was 0.35 mm.

16×10^{-6} g cm^{-2} (20 mg evaporated), while for thicker films, I_B approaches a constant value of about 40 percent I_0. I_B is a function of the apertures in the system and would vary from one microscope to another, but I_s should not vary greatly unless spherical aberration constants are very different.

III. TEST OF THE EXPONENTIAL LAW FOR I_T

In order to test Eq. (2), I_T must be distinguished from I_K and I_B at the image plane. Since I_T is represented by a practically plane wave, while I_K is represented by waves of appreciable aperture, the separation can be accomplished by means of a dark-field mask as indicated in Fig. 8. A 400-per-inch copper Lektromesh screen was placed in the cap of the specimen holder, followed by a brass spacer 0.35 mm thick and finally by the specimen on a 200- or 100-per-inch screen. Areas over the opaque portions of the mask appear in dark-field, imaged by that part of I_K passing through the opening in the mask, while areas over the openings are

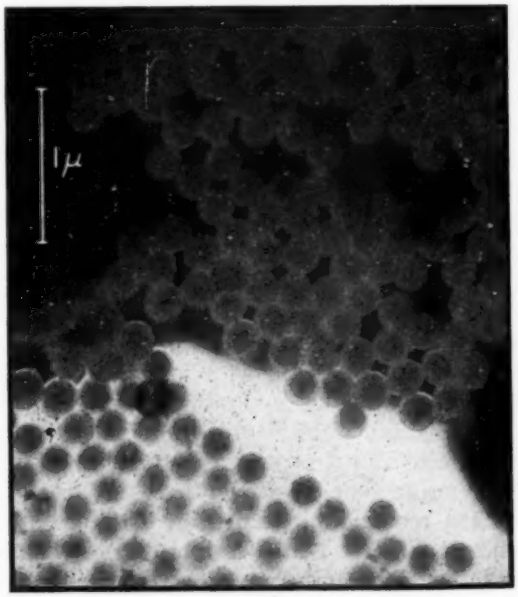


FIG. 9. Micrograph of polystyrene latex spheres on a collodion film, recorded at the region of transition from bright-field to dark-field as indicated in Fig. 8.

imaged with I_T and a part of I_K. The result is illustrated in Fig. 9 showing a micrograph of polystyrene latex particles on a collodion film, recorded at the edge of a dark-field mask. The unsharpness of the edge is due to the finite aperture of illumination. The arrangement shown in Fig. 9 is a more convenient method of dark-field observation than a similar system previously described,[3] particularly if the upper and lower grids are 100 and 400 mesh, respectively. If the distance between grids is increased, the intensity falls off less rapidly into the dark-field region. It is not necessary for dark-field (or bright-field) observation to position diaphragms at large distance from the object as has been the practice.

A typical tracing at the edge of a dark-field mask

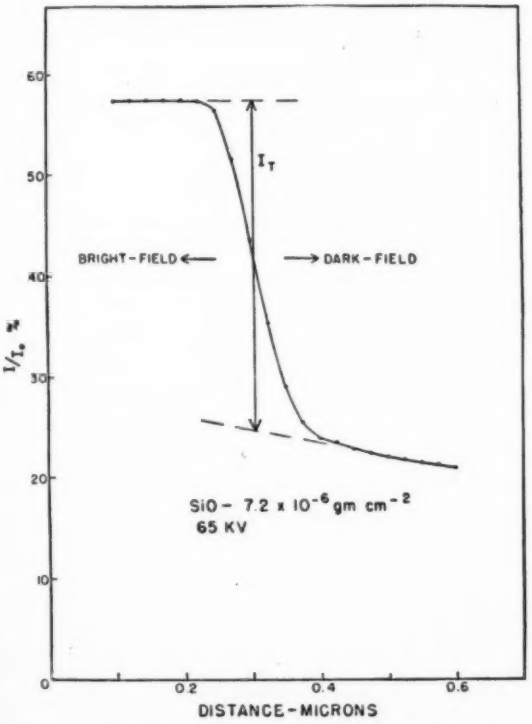


FIG. 10. Tracing at the edge of a dark-field mask illuminated through an SiO film, illustrating the method of measuring I_T/I_0.

below an SiO film is shown with the indicated graphical construction used to determine I_T in Fig. 10. At the edge, there is a sudden drop in intensity because of the subtraction of I_T, and just inside the edge the intensity should be about $\frac{1}{2}I_K$ plus background. Plates were recorded as for the measurements of I_c, and measurements were made directly from density plots. Resulting measurements of I_T are shown in Fig. 11, where $\log_e I_0/I_T$ is plotted as a function of the evaporated weight of SiO. Since the slope of the line is much greater than that in Fig. 6, the horizontal scale has been expanded by a factor of about 3, exaggerating the uncertainties in relative film thickness. The scatter of plotted points (each represents the average of three determinations) is attributed to experimental errors, and it is concluded that the form of Eq. (2) is verified within the limitations of the method. As for I_c, the range of the curve is surprisingly great, extending to ratios of I_T/I_0 less than 5 percent. When much thicker films are examined, the edge is invisible as the total intensity is almost the same on both sides. The slope of the line in Fig. 12, after conversion of abscissae to absolute values is $S_T = 13 \times 10^4$ cm^2 g^{-1}. Since $k = 1 - S_c/S_T$, we also have $k = 0.70$. This means that about 70 percent of the intensity lost in a layer, dw, by an initially plane wave occurs as coherent intensity within the effective aperture of the objective.

[3] C. E. Hall, J. Appl. Phys. 18, 588 (1947).

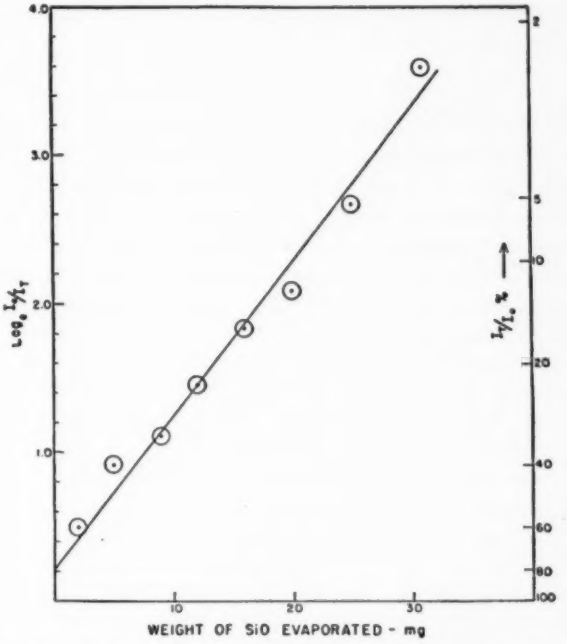


FIG. 11. Experimental test of the exponential law for I_T, Eq. (2).

IV. INTENSITY IN DARK-FIELD

It is not possible, generally, to predict the intensity in dark-field, since it depends on the geometry of the mask and the location of the point to be imaged, with respect to the opening. However, it is apparent that

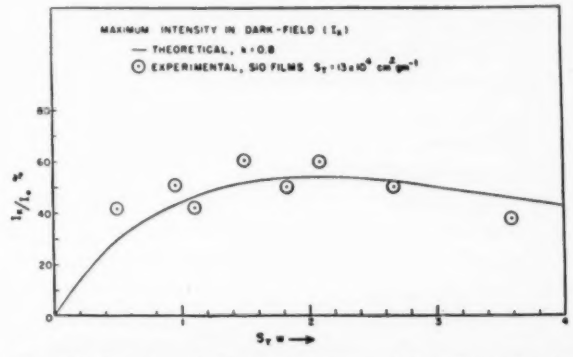


FIG. 12. Experimental test of the form of Eq. (7) for I_K, the maximum intensity available for dark-field imaging.

the maximum intensity available for dark-field imaging is I_K as given by Eq. (6), or

$$I_K = I_0 e^{-S_T w}[e^{kS_T w} - 1]. \quad (7)$$

The equation indicates that I_K rises from zero for $w=0$ to a maximum, and eventually decreases to zero for very large w. That the intensity in dark-field *versus* object thickness must follow such a trend was first deduced qualitatively by von Borries and Ruska.[4]

Just inside the edge of a dark-field mask the intensity should be approximately $\frac{1}{2}I_K$ plus background; and since the background in dark-field has been observed to be quite small with the masks used, I_K can be estimated from the set of traces as represented by Fig. 10. Although high accuracy cannot be expected because of the unsharpness of the transition region, I_K was taken as twice the lower ordinate of the line indicating I_T in Fig. 10; and the intensities so obtained were plotted as a function of $S_T w$ as shown in Fig. 12. The curve drawn in Fig. 12 was calculated from Eq. (7) using the measured value of S_T; but to obtain the best fit, k was assumed to be 0.8 rather than the measured value 0.7,

TABLE II. Electron scattering constants.
65 kv. $\lambda = 0.047$A.

Substance	Z	S_e 10^4 cm^2 g^{-1}	$a = S_e/\lambda^2$ 10^{23} g^{-1}	σ_e 10^{-18} cm^2
Be	4	3.0	1.4	0.45
SiO	14+8	3.7	1.7	1.4[a]
Cr	24	7.3	3.4	6.3
Ge	32	7.0	3.2	8.4
Pd	46	4.5	2.1	8.0
Pt	78	3.0	1.4	9.8
U	92	3.3	1.5	13.

[a] Average per atom. Also σ_T for SiO $= 4.7 \times 10^{-18}$ cm^2 average per atom.

which gives a curve of similar shape, a little below that shown. In consideration of the uncertainty in measuring I_K from traces such as that in Fig. 10, the results in Fig. 12 are in reasonable agreement with the theory.

V. SCATTERING CONSTANTS OF ELEMENTS

Measurements of the scattering constants S_e and S_T and the corresponding atomic cross sections are of interest, as they bear on theories of electron scattering and on the practical problems of quantitative densitometry in electron microscopy. Assuming the validity of the exponential relations (2) and (5), it is evident that S_e, S_T, and k can be obtained for any material from two measurements. The most difficult part of the procedure is that of obtaining an accurate value of w for the thin films required; and the major uncertainty is the possible dependence of S_e on state of chemical combination, mixture, or crystallinity. A few measurements have been made of S_e for elements selected at intervals through the periodic table; and, although the probable accuracy of the measurements is not high, the preliminary results are presented as an illustration of the methods and the problems to be met.

Films of the various elements were deposited by vacuum evaporation on SiO supports whose scattering power was known from previous experiments, and w was determined from the weight increment of cover slips (area about 5 cm) located beside the specimens during evaporation. The procedure was followed as for the SiO films, except for subtraction of the scattering power due to the SiO support. Results for Be, SiO, Cr, Ge, Pd, Pt, and U are listed in Table II. The tabulation shows a systematic increase with Z of σ_e, the total cross section for incoherent scattering as would be expected; but the results are too few to be used for the establishment of a theoretical or empirical relation between scattering constants and Z at this time.

Before measurements such as those in Table II can be extended to other elements, it will be advisable to examine some of the difficulties of the method. Consistency has been satisfactory in the densitometric part of the measurements and in results from different parts of a specimen or different specimens of the same material. Unfortunately, some unaccountable inconsistencies have appeared in several instances in checks made on the weight determinations. The weighing problem is not a simple one, since increments of the order of 100 μg must be measured accurately and the procedure assumes that the weight of a glass cover slip is constant between the first and second weighing, during which time it is heated to some extent in a vacuum chamber. Errors in this part of the procedure are more serious for the direct determination of S_T, where thinner films must be used. According to the theory, however, if S_e is known, it is not necessary to determine the weight of films used to find S_T. From two measurements as represented in Figs. 4 and 10 on a film of unknown thickness, we can find $(1-k)$, and then $S_T = (1-k)S_e$. Scattering constants for many elements can be measured only through the use of chemical compounds; and although we do not expect S_e to be critically dependent on the state of chemical combination, experimental evidence to support this opinion would be desirable.

VI. CONTRAST DUE TO SCATTERING

The perceptibility of an edge where scattering power changes abruptly may be enhanced in an electron micrograph by recording the image slightly out of focus so as to produce interference fringes. We are concerned here, however, only with differences in intensity due to scattering power. Essentially, the problem is that of detecting a difference in intensity, ΔI_e, against a field of intensity $I_e + I_B$. Obviously, it is desirable to reduce I_B by suitable apertures so that $I_e \gg I_B$. This may not always be possible with very thick films, since a part of the wave producing I_B is contained within α_0. If with a very thick film the diameter of the physical aperture is reduced to α_0, a part of I_K will be removed because

[4] B. von Borries and E. Ruska, Z. Physik 116, 249 (1940).

it is not a sharply defined pencil and S_e will be increased accordingly. If electrostatic interaction between the beam and the aperture is neglected, I_B under such circumstances reduces to that part of the beam which has been changed appreciably in wavelength, but still passes through the system. A complete analysis requires a knowledge of the angular distribution of I_K, I_B and a knowledge of the velocity distribution of the inelastic component, as well as its angular distribution. Such an analysis cannot be undertaken here, but it is interesting to estimate the minimum observable difference in mass thickness subject to the simplifying assumption that $I_e \gg I_B$.

In the estimation of the minimum detectable mass thickness, two methods of observation are to be considered: (1) a photographic record observed by eye and (2) a photographic record observed with a microdensitometer. In the first case it is desirable to expose the plate so that I_e is on the straight-line portion of the $H-D$ curve and then,

$$\Delta D=\gamma(\Delta I_e/I_e)\log_{10}e=-\gamma S_e\Delta w\log_{10}e. \quad (8)$$

If the plate is observed by transmission and we assume the contrast sensitivity of the eye under practical viewing conditions is about $0.02=\Delta T/T$, where T is the transmission, we have

$$\Delta D=-0.02\log_{10}e \quad (9)$$

and

$$\Delta w=0.02/\gamma S_e. \quad (10)$$

For densitometric observation we can assume $D\propto I_e$ up to about $D=1$ to a fair approximation, as indicated in Fig. 3; and, therefore,

$$\Delta D/D=\Delta I_e/I_e=-S_e\Delta w. \quad (11)$$

The scale reading of a microdensitometer is usually linear with transmission rather than density, so that $\Delta D/D$ varies across the scale. A density difference of 0.01 at $D=1$ was the minimum significant increment on the instrument used for this work; and, therefore,

$$\Delta w=0.01/S_e, \quad (12)$$

which is the same as for the eye with the contrast sensitivity assumed when $\gamma=2$. Whether such a small density difference is regarded as significant in practical circumstances will depend also on the abruptness of the change at an edge, that is to say, on the resolution.

Consider a film consisting of light elements such that $S_e=4\times10^4$ cm^2/g and density $\rho=1$ g cm^{-3}. Either Eq. (10) or Eq. (12) then indicates a minimum thickness, $\Delta w/\rho$, of about 25A for 65 kv, at which S_e was measured. Consider also the example of uranium with $S_e=3.3\times10^4$ cm^2/g and $\rho=18.7$ g/cm^2. It is readily calculated that $\Delta w/\rho=1.5$A. Note that neither Eq. (10) nor Eq. (12) contains the film thickness, suggesting that the minimum detectable w is unaltered for very thick specimens. This might appear to be in contradiction with experience, but it is the natural consequence of assuming $I_B \ll I_e$. Actually, if the background can be reduced to a low value, the minimum detectable difference in scattering power is largely independent of total scattering power except to the extent that deterioration in resolution alters the abruptness of an edge and affects the judgment of what constitutes a significant difference in density. The notion that structures are invisible through thick parts of a specimen because of lack of "penetration" is erroneous. The difficulty is almost always that $I_B > I_e$ in such regions, and visibility can be restored even at moderate beam potentials by appropriate aperturing.[5] If the aperture is reduced to considerably less than α_0, S_e must increase, resulting in a decrease of Δw.

VII. EFFECT OF BEAM POTENTIAL

Although no measurements have been made of the effect of beam potential on the scattering constants, previous theories[6,7] have been consistent in predicting that total elastic and inelastic cross sections vary as λ^2. In the absence of measurements it is reasonable to assume that $S_T\propto\lambda^2$ and to set $S_e=a\lambda^2$, where a is a constant as listed in Table II. Consider the problem of detecting a thin particle of scattering power $a_2\lambda^2 w_2$ on a thick film of scattering power $a_1\lambda^2 w_1$. Then we have

$$\Delta I_e=I_0\exp(-a_1\lambda^2 w_1)[1-\exp(-a_2\lambda^2 w_2)]. \quad (13)$$

As was shown in the previous section, a small difference in intensity is most perceptible when the density is of the order of 1 to 1.5, so that I_e+I_B is fixed and the problem reduces to finding λ such that ΔI_e is a maximum. Differentiating Eq. (13) with respect to λ and setting the result equal to zero, we obtain, finally,

$$\lambda_{opt}^2=\log_e[(a_2w_2/a_1w_1)+1]/a_2w_2. \quad (14)$$

If now $a_2w_2 \ll a_1w_1$, Eq. (14) become

$$\lambda_{opt}^2\doteqdot 1/a_1w_1, \quad (15)$$

indicating that the optimum voltage for the observation of small particles on a thick substrate is independent of the nature of the particles and is a function only of the scattering power of the substrate. If relativity is neglected, Eq. (15) becomes

$$V_{opt}\doteqdot a_1w_1\times150\times10^{-16}\text{ volts}. \quad (16)$$

Consider a film of low atomic number, 1000A thick, density 1.5, and $a=1.7\times10^{23}$ (as for SiO). Then, $V_{opt}=38{,}000$ volts. In the light of experience the value is reasonable, but it is somewhat lower than the 50-kv accelerating potential of most microscopes used in this country. There are many factors in addition to contrast to be considered in the choice of a suitable operating potential, such as variability in specimens, heating

[5] J. Hillier, J. Bact. **57**, 313 (1949).
[6] L. Marton and L. I. Schiff, J. Appl. Phys. **12**, 759 (1941).
[7] B. von Borries, Z. Naturforsch. **4a**, 51 (1949).

effects, contamination, effect of disturbing fields, intensity, and others. However, there is a tendency to operate at voltages up to about 100 kv, and it is interesting to consider the effect of beam potentials greater than V_{opt}.

With beam potentials greater than the optimum for contrast by scattering we expect:

(1) An increase in I_e/I_B, since S_e decreases and the effect is the same as moving to lower abscissas in Fig. 7. Thus, if the apertures are not effective in decreasing the background sufficiently with respect to I_e, structures in thick specimens will become visible, which, however, could be seen without raising the voltage if the background were removed.

(2) An increase in I_T/I_e. From Eqs. (2) and (5)

$$I_T/I_e=\exp(-ka\lambda^2w/1-k), \tag{17}$$

and also at V_{opt}, a $\lambda^2w=1$. Assume, for example, $k=0.7$ as for the SiO films; and it is found that at the optimum potential, $I_T/I_e=0.1$, which is undesirably small, since I_T is the component most capable of producing a high resolution image. The remainder of I_e is I_K producing images in quality like those in dark-field. Equation (17) can also be written

$$I_T/I_e=\exp[-kV_{opt}/(1-k)V]. \tag{18}$$

Increasing the voltage above V_{opt} increases the ratio I_T/I_e in greater proportion. For example, with $k=0.7$ as before, doubling the voltage raises I_T/I_e from 10 percent to 30 percent.

(3) The minimum detectable mass thickness will be larger, since $S_e\propto 1/V$ (neglecting relativity) and Eqs. (10) and (12) then show that

$$\Delta w\propto V. \tag{19}$$

Note that the numerical estimates of Δw made in Sec. VI apply only at 65 kv, or considerably above V_{opt} for most specimens.

VIII. CONCLUDING REMARKS

There are, as yet, insufficient data on the scattering constants of the elements to permit an outline of the possibilities of quantitative densitometry in image interpretation. Within the limitations of the theory and experimental methods, Eqs. (2), (5), and (6) suggest numerous relations, measurable from photographic records, according to such knowledge of scattering constants and specimen composition as can be brought to bear. If no constants are known, the exponential form of I_e lends itself readily to the measurement of relative amounts of any single substance on the same substrate. In any case, the main precaution to be exercised in the measurement of the usual bright-field image intensities is that of determining I_B or insuring that it is of negligible proportions.

Experimental data on scattering phenomena directly applicable to the problems of electron microscopy are scarce, although a number of theoretical papers bearing on the problem have been presented, most of which have been summarized and discussed by von Borries.[7] Inelastic scattering, according to formulas derived by Koppe,[8] is mostly confined to angles within the effective apertures of electron microscope objectives. This conclusion is supported by experimental results obtained by Hillier and Ramberg,[9] who also find that the most probable velocity losses are so small that resolution is not seriously affected by the inelastic component. Elastic scattering appears to be the main contribution to intensity at large angles. Of the several available formulas for the angular distribution of the elastically scattered intensity, von Borries favors one derived by Molière,[10] which formula, in conjunction with others, has been used by Hillier and Ramberg[11] for the calculation of scattering cross sections at 50 kv of several elements. Their calculations may not be compared directly with the experimental results in Table II, since the elements are different. It is interesting to note, however, that for the effective aperture of the lens in the present study, their calculated cross sections are of the same order of magnitude, though generally a little greater than those in Table II.

Scattering formulas, such as those referred to, embody the concept of electron deflections by single atoms and are usually applied with the assumption that the resultant intensity caused by groups of atoms is the arithmetic sum from individual atoms. If, however, ordered or semi-ordered regions exist, an adequate description of the transmitted intensity must take account of the coherency of scattering emanating from the constituent atoms, an effect most evident when Bragg reflections occur. Some of the effects of ordered structures on the distribution of transmitted intensity have been discussed by Boersch.[12] Since few substances can be considered as completely amorphous, some modification of the formulas based on monatomic scattering may be necessary in predicting effective scattering cross sections applicable to solids.

---

[8] H. Koppe, Z. Physik **124**, 658 (1948).
[9] E. G. Ramberg and J. Hillier, J. Appl. Phys. **19**, 678 (1948).
[10] G. Molière, Z. Naturforsch. **2a**, 133 (1947).
[11] J. Hillier and E. G. Ramberg, Z. angew. Physik **2**, 19 (1950).
[12] H. Boersch, Z. Naturforsch. **2a**, 615 (1947).

X-Ray Measurement of Long-Range Order in β-AgZn*

L. MULDAWER†
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received February 21, 1951)

The replacement of small amounts of Ag by Au in AgZn raises the critical temperature for long-range ordering and completely suppresses the zeta-phase. Measurements of T_c were made in a high temperature camera using several percentages of Au. Extrapolation to zero content of Au gives $T_c = 272°C$ for pure AgZn. The variation of long-range order with temperature was determined from measurements of the integrated intensity of the (111) superstructure line, using a compact of filings of composition 3.5 Au, 46.5 Ag, and 50.0 Zn. The variation is similar to that found in β-brass, and is in agreement with the predictions of the Cowley theory.

I. INTRODUCTION

THE β-phase of the Ag-Zn system exists in a narrow range near the composition AgZn. Above about 280°C, β-AgZn has a disordered body-centered cubic structure.[1,2] Quenching from above 280°C produces an ordered CsCl structure. Slow cooling from above 280°C, or raising the temperature of an ordered sample to the range 100°–280°C produces a complex hexagonal structure known as the zeta-phase. For β-AgZn, it is not possible to measure the temperature T_c, at which long-range order sets in, or to measure the variation of long-range order with temperature as was done with β-CuZn.[3] Presumably, T_c is somewhere below 280°C, and the interesting temperature range, where long range order sets in, is completely masked by the more stable zeta-phase.

In the Au-Zn system there is a β-phase at about the composition AuZn, which shows only an ordered CsCl structure at all temperatures studied and presumably up to the melting point.[1] Since T_c is high for AuZn, the idea suggests itself that by adding a small amount of AuZn to AgZn it should be possible to raise T_c to above the transition temperature to the zeta-phase, and thus give a temperature range in which the ordered cubic form would be stable. The replacement of Ag by Au in AgZn might also increase sufficiently the stability of the ordered cubic form to eliminate the zeta-phase altogether. The experiments described here indicate that the replacement of Ag by Au raises T_c by an amount which increases with the Au content, and for 3 percent or more of Au, eliminates completely the zeta-phase.

By using samples with various amounts of Au, it is now possible to measure T_c and the variation of long-range order with temperature exactly as with β-CuZn. Extrapolation of the results to zero content of Au will then give T_c and the long-range order variation for the ordered cubic form of AgZn.

TABLE I. Atomic percentage compositions.

Item	Au	Ag	Zn
1	2.4	47.6	50.0
2	3.5	46.5	50.0
3	5.0	45.0	50.0

II. MEASUREMENTS OF THE CRITICAL TEMPERATURE

Alloy specimens were made with the compositions given in Table I.

Only the first showed some evidence of zeta-phase formation when the filings were annealed at 200°C; however, it was still possible to measure T_c. Patterns were made with filtered CuKα, using cylindrical samples of filings in an evacuated high temperature camera. Patterns were run for each sample at a number of closely spaced temperatures. An estimate of the temperature at which the intensity of the superstructure lines approaches zero gives the values of T_c, which are plotted in Fig. 1. An extrapolation of these values to zero content of Au gives a value $T_c = 272°C$ for AgZn.

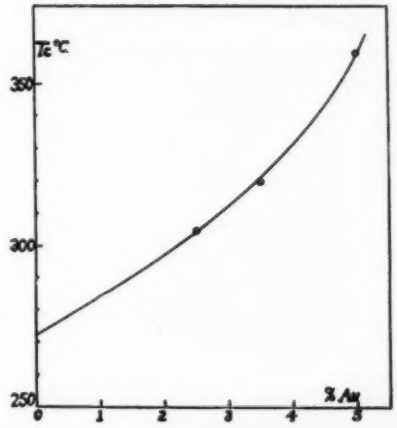


FIG. 1. Critical temperature for long-range ordering as a function of the atomic percent of gold in the alloy (x) Au, (50-x) Ag, 50 Zn.

* Research sponsored by the ONR, under contract N5 ori-07832.
† Work done at MIT in summer of 1950. Regular address: Temple University, Philadelphia, Pennsylvania.
[1] E. A. Owens and I. G. Edmunds, Proc. Phys. Soc. (London) 50, 389 (1938).
[2] Andrews, Davies, Hume-Rothery, and Oswin, Proc. Roy. Soc. (London) A177, 149 (1941).
[3] D. Chipman and B. E. Warren, J. Appl. Phys. 21, 696 (1950).

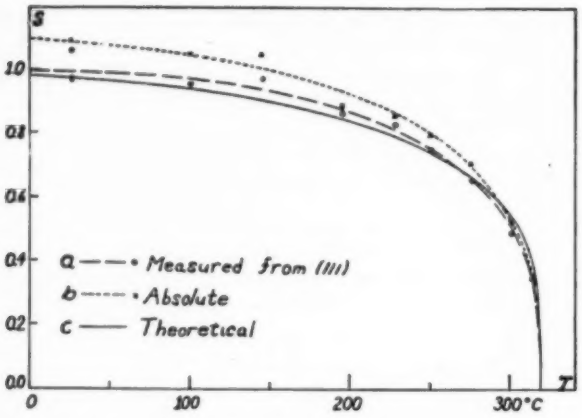


FIG. 2. Long-range order as a function of temperature for the alloy 3.5 Au, 46.5 Ag, 50 Zn. (a) Relative values of S adjusted to $S=1.0$ at 0°C. (b) Absolute values of S obtained by comparison with the (200) reflection. (c) Theoretical Cowley curve for $T_c=318$°C.

III. MEASUREMENT OF LONG-RANGE ORDER

The measurement of long-range order was made with a compact of filings of atomic composition 3.5 Au, 46.5 Ag, 50 Zn. The compact was mounted in a furnace provided with thin aluminum windows and centered on a Geiger counter spectrometer. A stream of hydrogen through the furnace prevented oxidation. Temperatures were measured by a thermocouple soldered to the plate supporting the sample. A bent quartz crystal was used as a focusing monochromator for the CuKα-radiation.

Measurements of the integrated intensity were made at various temperatures for the fundamental line (200) and the superstructure line (111). The region around each reflection was first investigated, using a fine slit at the Geiger counter to determine peak positions, peak widths, and flatness of background. The counter slit was then opened to subtend 2° in 2θ, and a measurement was made which included the whole line intensity, followed by a similar measurement of background only on either side of the peak. The average of the two background readings subtracted from the peak reading gave the integrated intensity of the line.

The measurements of (200) were used to determine a temperature factor. For the conditions of this investigation, the temperature correction is small and the factor is adequately represented by

$$TF=\exp[-\alpha(h^2+k^2+l^2)T].$$

A plot of the logarithm of the (200) intensity against temperature gave $\alpha=1.75\times10^{-4}$ per degree. This value was used for a temperature correction for the (111) reflection.

For partially ordered β-AgZn the structure factor has the form

(111) Superstructure line $F=(f_{Ag}{}^*-f_{Zn})S$,
(200) Fundamental line $F=(f_{Ag}{}^*+f_{Zn})$,

where $f_{Ag}{}^*$ is a weighted scattering factor of Ag and Au, and S is the Bragg and Williams long-range order parameter.[4] The integrated intensity of the superstructure line (111) properly corrected for the temperature effect is proportional to S^2. The square root of the corrected intensity is then proportional to S. If we assume perfect order at room temperature, the plot of the square root of the corrected (111) intensity can be normalized to $S=1$ at room temperature. Based on this assumption, the variation of S with T for AgZn is represented by Fig. 2a.

The integrated intensity of either the (111) or the (200) line depends upon

$$P\sim m(LP)(TF)F^2,$$

where m is the multiplicity, (LP) is the Lorentz polarization factor, (TF) is the temperature factor, and F is the structure factor. The ratio of the integrated intensities of the (111) and the (200) reflections gives directly an absolute value of S. The values obtained in this way are shown by Fig. 2b. Since values of S greater than unity are not possible, there is some error in the curve. The error might be due to neglect of extinction, since extinction would be greater for (200) than for (111), and this would make the measured value of S come out too high. A small error in the proper Hönl correction to the atomic scattering factors could also give values of S which are too high.

Figure 2a, obtained by arbitrarily setting $S=1.0$ at room temperature, probably gives the best representation of S as a function of temperature. For a β-brass type structure Cowley's theory[5] predicts a variation of long-range order with temperature of the following form.

$$\ln(1+S^2/1-S^2)=2(T_c/T)S^2.$$

Using $T_c=318$°C, the computed variation of S with T is given by Fig. 2c. The agreement between curves a and c is well within the experimental error of the measured curve and the approximations of the theoretical curve

The variation of long-range order with temperature, which is represented by Fig. 2a, relates to AgZn containing 3.5 percent Au. Similar curves for two other samples with different Au content would allow an extrapolation to zero Au content. It was not felt that this additional work was justified since the result would not differ significantly from that obtained by treating the curve 2a as a function of T/T_c and changing T_c to the value 272°C, corresponding to pure AgZn.

The addition of small amounts of Au to AgZn gives an alloy in which the order-disorder transformation is similar to that occurring in β-brass and in good agreement with the predictions of the Cowley theory. Since the addition of only small amounts of Au are necessary

[4] W. L. Bragg and E. J. Williams, Proc. Roy. Soc. (London) 145, 699 (1934).
[5] J. M. Cowley, Phys. Rev. 77, 669 (1950).

to keep AgZn in the β-form, it is clear that the energies of the beta- and zeta-structures are very close for pure AgZn.

The author wishes to thank Professor B. E. Warren who provided the facilities for this research and whose advice aided in the course of the work, and Professor B. L. Averbach for advice in the preparation of the alloy samples.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

A New X-Ray Diffraction Method for Studying Imperfections of Crystal Structure in Polycrystalline Specimens*

ALFRED J. REIS, JEROME J. SLADE, JR., AND SIGMUND WEISSMANN†
Materials Research Laboratory, Rutgers University, College of Engineering, New Brunswick, New Jersey and the Department of Chemistry of the Polytechnic Institute of Brooklyn

(Received November 15, 1950)

A new double spectrometer method for the study of polycrystalline materials is developed. From intensity data of the diffracted spots and known angular settings of the specimen, the angular range of reflection of the misaligned scattering regions in a crystallite is determined statistically.

The instrument and the technique of the multiple exposure method is described. The statistical evaluation of the data is discussed and a mathematical treatment introducing correction terms as functions of the azimuth and of the Bragg angle is presented.

A commercial silicon powder specimen was investigated. The average angular range of reflection was found to be nearly the same for all (*hkl*)-reflections with exception of (111), (333), and (444).

The study of broadening effects with the same instrument has been started.

The new double spectrometer method for polycrystalline specimens described in this paper is based on the examination of numerous microscopic spots reflected by individual crystallites. The small spots on the diagram facilitate the detection and measurement of small broadening effects. Independently, the pattern of angular misalignment of the same reflecting crystallite can be accurately determined. For both broadening and misalignment, the method offers statistical information which is representative of the specimen and which can be compared for different (*hkl*)-reflections.

The total information on imperfections obtained in this manner may serve to correlate experimental observation with the theory of dislocations and hence, may be useful in the understanding of structure-sensitive properties of solids.

THEORY OF THE NEW METHOD

THIS double spectrometer method is based on the irradiation of the polycrystalline specimen with x-rays of small horizontal convergence.[1] (As the crystallites are much smaller than the focal spot, it is the convergence, not the divergence which matters for the angular range of irradiation.) This is achieved by reflecting the primary x-ray beam from a fairly perfect monocrystal, which is adjusted in vertical position on the first spectrometer axis. Although the vertical convergence K of the beam is not altered by the reflection from the monocrystal, the horizontal convergence has thus been reduced to a range L, which is determined by the perfection of the monocrystal. The reflected beam is directed toward the second axis of the instrument on which the test specimen is mounted.

In a stationary exposure some crystallites of the specimen are in reflecting positions; from each of these, a reflected beam originates. On the cylindrical film, the beams are recorded as microscopic spots which are arranged along Debye-Scherrer curves. The stationary spot diagram is recorded several times on the same film, and between exposures, small discrete changes are made of both the angular setting of the specimen and the position of the film. By changing the setting of the goniometer axis the crystallites in the specimen are stepwise rotated through the range of reflecting positions. Because of the filmshift, a separate spot is recorded for a crystallite at each setting within the range of its reflecting positions. As a result of the procedure, the multiple exposure with different specimen settings gives rise to an array of spots on the film.

The length of the array can serve as an indication of the range of reflecting positions. However, more significant data for the study of angular misalignment in the crystallites are obtained by observing how the intensity of the spot varies within an array.

The shape and characteristics of the diffracted spots within an array are a visual manifestation of the penetration of the reciprocal lattice sphere with the sphere of reflection. The spot shape in the diagram varies according to the extension of the contact of the reciprocal lattice sphere with the sphere of reflection. This accounts for the change of shape and intensity of the diffracted spots in the sequence of the array. If the intensity is plotted *versus* angular setting of the specimen, one obtains a distribution curve, and the half-width (the width extending from one half-maximum to the other) is an indication of the angular range of reflection of the crystallite.

* Supported by the ONR under contract ONR-454 Task No. 1.
† This paper is abstracted from a dissertation of S. Weissmann in partial fulfillment of the requirements for the degree of doctor of philosophy of the Polytechnic Institute of Brooklyn.
[1] A. Reis, Compt. rend. 205, 369 (1937).

If one observes all the arrays of spots along a Debye-Scherrer curve, one will discover that the length of an array as well as the half-width of the distribution curve increase in extent, as one moves away from the equatorial plane upward or downward. This increase represents, however, an apparent increase of the angular range of reflection and is mainly due to two factors. The first factor that will cause an apparent increase of the angular range of reflection is the velocity factor.[2] The second factor is due to the vertical convergence of the beam, which remains unaltered after reflection from the monocrystal. Hence, in the evaluation of the intensity distribution of the array of spots, these two factors must be considered and proper correction terms introduced. The subsequent treatment leads to mathematical expressions of the angular range of reflection, taking into account the corrections mentioned above.

Let the center of coordinates be at the center of the specimen, the $x-y$ plane coinciding with the equatorial plane, the x-axis coinciding with the central ray of the irradiating beam.

If the angle of convergence K of the irradiating beam and the angular imperfection L of the parallelizing crystal are small, in the sense that $\sin K=K$, $\sin L=L$, $\cos K=\cos L=1$ when K and L are expressed in radians, then the unit vector

$$r_0=(1,\eta,\xi);\quad -K/2\leq\xi\leq K/2;\quad -L/2\leq\eta\leq L/2 \quad (1)$$

represents the solid pencil of rays with rectangular base converging to the specimen from the parallelizing crystal. The curvature of the bounding surfaces of this pencil vanishes with K^2.

Let θ be a Bragg angle and n_0 the unit normal of the reflecting plane of a crystallite. A ray specified by fixed values of ξ, η will be reflected whenever

$$r_0\cdot n_0=-\sin\theta. \quad (2)$$

This is the equation of a cone which we shall call the conjugate cone.

The family of vectors representing the Debye-Scherrer cone is

$$p_0=r_0+2\sin\theta n_0 \quad (3)$$

provided n_0 is determined by Eq. (2).

As ξ, η are allowed to vary through all permissible values, Eq. (2) represents a family of cones that intersects the unit sphere in a zone. A crystallite plane will be in reflecting position whenever its normal pierces this zone.

If $n_0=(l,m,n)$, $l=\cos\psi\cos\phi$, $m=\cos\psi\sin\phi$, and $n=\sin\psi$, where ψ is the latitude measured from the equatorial plane and ϕ the longitude measured from the x-axis, then the parametric equation of the reflecting zone is

$$\begin{aligned} r_0\cdot n_0+\sin\theta&=l+\eta m+\xi n+\sin\theta\\ &=\cos\psi\cos\phi+\eta\cos\psi\sin\phi+\xi\sin\psi+\sin\theta=0\\ &\quad -K/2\leq\xi\leq K/2;\quad -L/2\leq\eta\leq L/2. \end{aligned} \quad (4)$$

Let ϵ be the angle, measured from the center of the unit sphere along a circle of latitude, subtended by a region of the sphere pierced by the misaligned normals of the crystallite. Then

$$\Delta\phi_1=\epsilon/\cos\psi \quad (5)$$

is the angle required to turn this region past a fixed longitude.

Let $\Delta\phi_2$ be the angle required to turn any point of the region of misalignment through the reflecting zone. If ϕ is measured from the center of the reflecting region, then, for azimuths $|\psi|=K/2$, reflection starts at $\phi-\Delta\phi_2/2$, where $\xi=-K/2$, $\eta=-L/2$, and ends at $\phi+\Delta\phi_2/2$, where $\xi=K/2$, $\eta=L/2$.

Substituting these values in Eq. (4) and subtracting the resulting equations, we obtain

$$\cos\psi\left\{\cos\left(\phi-\frac{\Delta\phi_2}{2}\right)-\cos\left(\phi+\frac{\Delta\phi_2}{2}\right)-\frac{L}{2}\left[\sin\left(\phi-\frac{\Delta\phi_2}{2}\right)+\sin\left(\phi+\frac{\Delta\phi_2}{2}\right)\right]\right\}=K\sin\psi.$$

This reduces to

$$\begin{aligned} 2\sin\frac{\Delta\phi_2}{2}-L\cos\frac{\Delta\phi_2}{2}&=\frac{K\tan\psi}{\sin\phi}\\ &=\frac{K\sin\psi}{(\cos^2\psi-\sin^2\theta)^{\frac{1}{2}}},\quad \text{where}\quad |\psi|>K/2. \end{aligned} \quad (6)$$

Since at the center of the reflecting region $\xi=\eta=0$ and

$$\begin{aligned} \cos\phi&=-\sin\theta/\cos\psi\\ \sin\phi&=(\cos^2\psi-\sin^2\theta)^{\frac{1}{2}}/\cos\psi \end{aligned}$$

when $\Delta\phi_2$ is small, Eq. (6) becomes

$$\Delta\phi_2=\frac{K\sin\psi}{(\cos^2\psi-\sin^2\theta)^{\frac{1}{2}}}+L \quad (7)$$

without this restriction $|\psi|>K/2$.

In this formula K and L may be expressed in degrees.

The angle through which reflection is observed is $\Delta\phi=\Delta\phi_1+\Delta\phi_2$. Using Eqs. (5) and (7), we finally obtain

$$\epsilon=\Delta\phi\cos\psi-\left(\frac{K\sin\psi}{(\cos^2\psi-\sin^2\theta)^{\frac{1}{2}}}+L\right)\cos\psi. \quad (8)$$

This formula, then, yields the angle of misalignment ϵ in terms of the total angle $\Delta\phi$ through which reflection is observed, the azimuth of the crystallite, the con-

[2] E. G. Cox and W. F. B. Shaw, Proc. Roy, Soc. (London) **A127**, 71 (1930).

vergence K of the irradiating beam, the angular imperfection L of the parallelizing crystal, and the Bragg angle θ.

It should be noted that Eq. (8) is strictly valid only for an angular range of reflection with well-marked boundaries. Since the observed boundaries are rather ill determined, we have taken $\Delta\phi = H'$ to be the angular displacement between the half-peak photometric intensities and used $0.67\ \Delta\phi_2$ instead of $\Delta\phi_2$, thus obtaining

$$\epsilon = H' \cos\psi - Q, \tag{9}$$

where

$$Q = 0.67\left[\frac{K \sin\psi}{(\cos^2\psi - \sin^2\theta)^{\frac{1}{2}}} + L\right]\cos\psi.$$

The theory of the photometric transformation of the spatial point density distribution of the misaligned normals has been developed and will be presented in a forthcoming paper.

To obtain the azimuth ψ in terms of the coordinates of the center of a spot in the Debye-Scherrer photograph, we use Eq. (3) with $\xi = 0$. If ϕ, ψ are the polar coordinates of p_0, we find that the z-direction cosine of p_0 is

$$\sin\psi_1 = 2 \sin\theta \sin\psi. \tag{10}$$

If y is the vertical distance from the equatorial plane on the film and R is the radius of the camera, then

$$\tan\psi_1 = Y/R,$$

$$\sin\psi_1 = \frac{Y/R}{[1+(Y/R)^2]^{\frac{1}{2}}}$$

and

$$\sin\psi = \frac{Y/R}{2 \sin\theta[1+(Y/R)^2]^{\frac{1}{2}}}. \tag{11}$$

DESCRIPTION OF THE DOUBLE SPECTROMETER

The double spectrometer is an instrument of great accuracy which can be easily attached to any commercial x-ray unit. The path of the emerging x-ray beam can be visualized by inspection of Fig. 1, which depicts a top view of the double spectrometer aligned to the x-ray window of a North American Philips diffraction unit. The x-ray beam passes through a collimator, B, which merely functions as a protection against stray radiation, and is then parallelized by a monocrystal which is mounted on a goniometer head E. The reflected beam then passes through a spectrographic slit system F and enters the cylindrical camera I. In the cylindrical camera the x-ray beam irradiates the polycrystalline specimen which is mounted on the goniometer head J in the center of the camera. The trap L fastened to the base plate, serves to block out the undeviated primary beam, while the scattered x-rays are recorded on the x-ray film K. Through the aid of a fine differential micrometer screw H, the angular

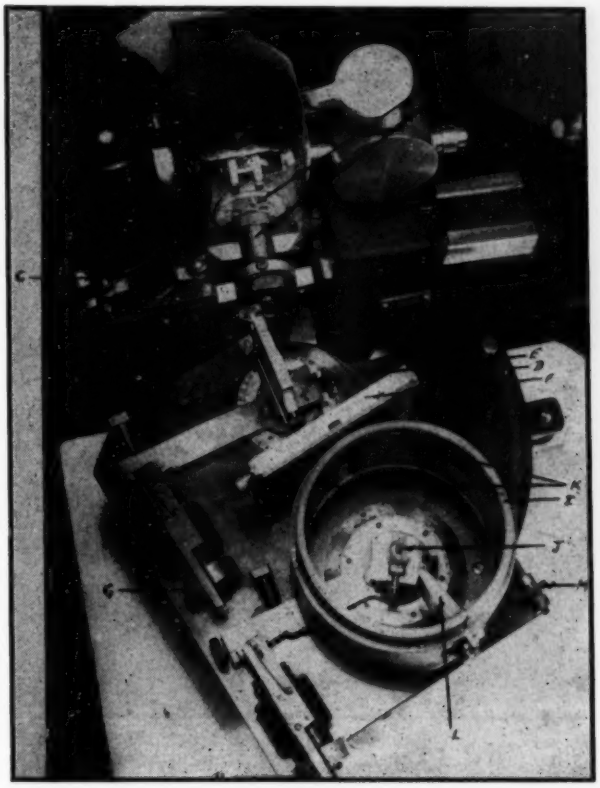


FIG. 1. Topview of the double spectrometer. A—Window of the x-ray tube of the North American Philips commercial unit. B—Collimator. C—Thermostatic control device. D—Upper movable platform. E—Goniometer head with monocrystal holder. F—Top of spectrographic slit. G—Differential micrometer screw for setting of monocrystal. H—Differential micrometer screw for setting of the polycrystalline specimen. I—Cylindrical camera. J—Goniometer head with holder of polycrystalline specimen. K—Location of x-ray film. L—Trap for primary beam. M—Screw arrangement for filmshift.

specimen settings can be effected. This fine differential micrometer screw is capable of an angular precision setting of ± 1 second of an arc. The filmshifts between single exposures can be obtained by means of the screw arrangement M. By virtue of the fine differential micrometer screw G, it is possible to impart to the monocrystal an angular rotation around the vertical axis of the goniometer head E with a precision of ± 1 second. One of the most conspicuous and highly useful features of the double spectrometer is the movable upper platform D, which can be rotated around the vertical axis of the goniometer head of the monocrystal. On account of the sturdy, immobile base plate, which provides for rigid support, this operation permits quick and easy alignment of the reflecting crystal and the polycrystalline specimen without disturbing the adjustment of the entire instrument with respect to the x-ray tube.

A cylindrical screen which is provided with a slot and which surrounds the polycrystalline specimen can also be used. This arrangement permits exposures of the equatorial zone only; and as the filmholder can

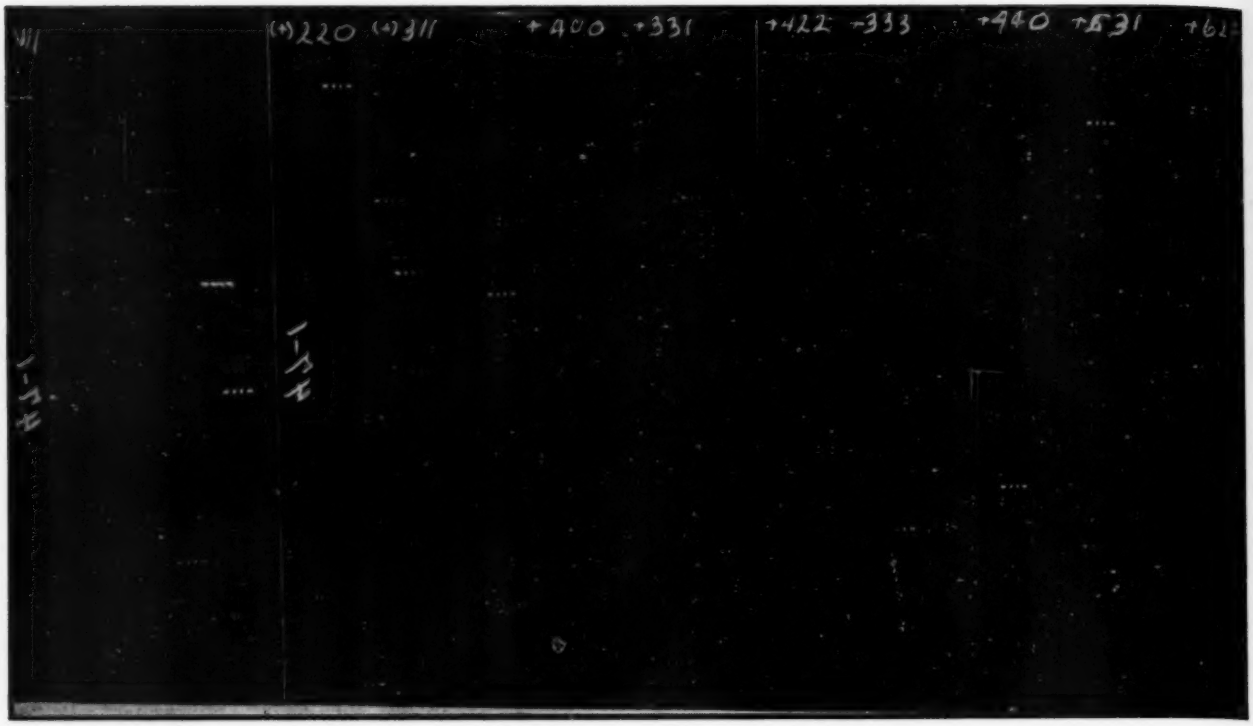


FIG. 2. Experiment 47-M. Multiple exposure of silicon powder, plus wing. Contact print. 4 specimen settings, 3 minutes change between settings. Many arrays of spots extend through the range of 9 minutes.

be screwed to various heights, it is possible to effect several exposures of the equatorial zone at various film heights.

For some crystallites the range of reflecting positions may fall partly within and partly beyond the total range of settings of a multiple exposure. In this case, the angular range of reflection of the crystallites will overlap the actual range of settings, and the array of spots will remain incomplete. In order to distinguish an incomplete array of spots from a small but complete

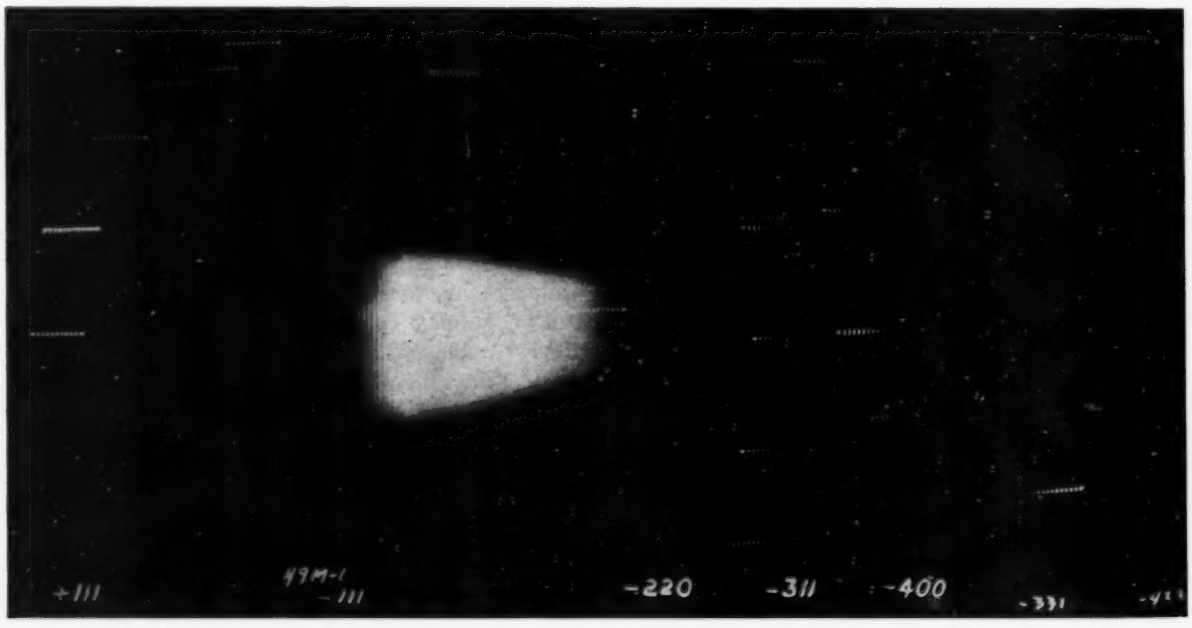


FIG. 3. Experiment 49-M. Multiple exposure of silicon powder. Enlargement 2.0×, 13 specimen settings, 1 minute change between the intermediate settings, 3 minutes change next to the terminal settings. The parallel lines between both (111) lines are the traces of one edge of the trap, and their distances measure the filmshift between single exposures.

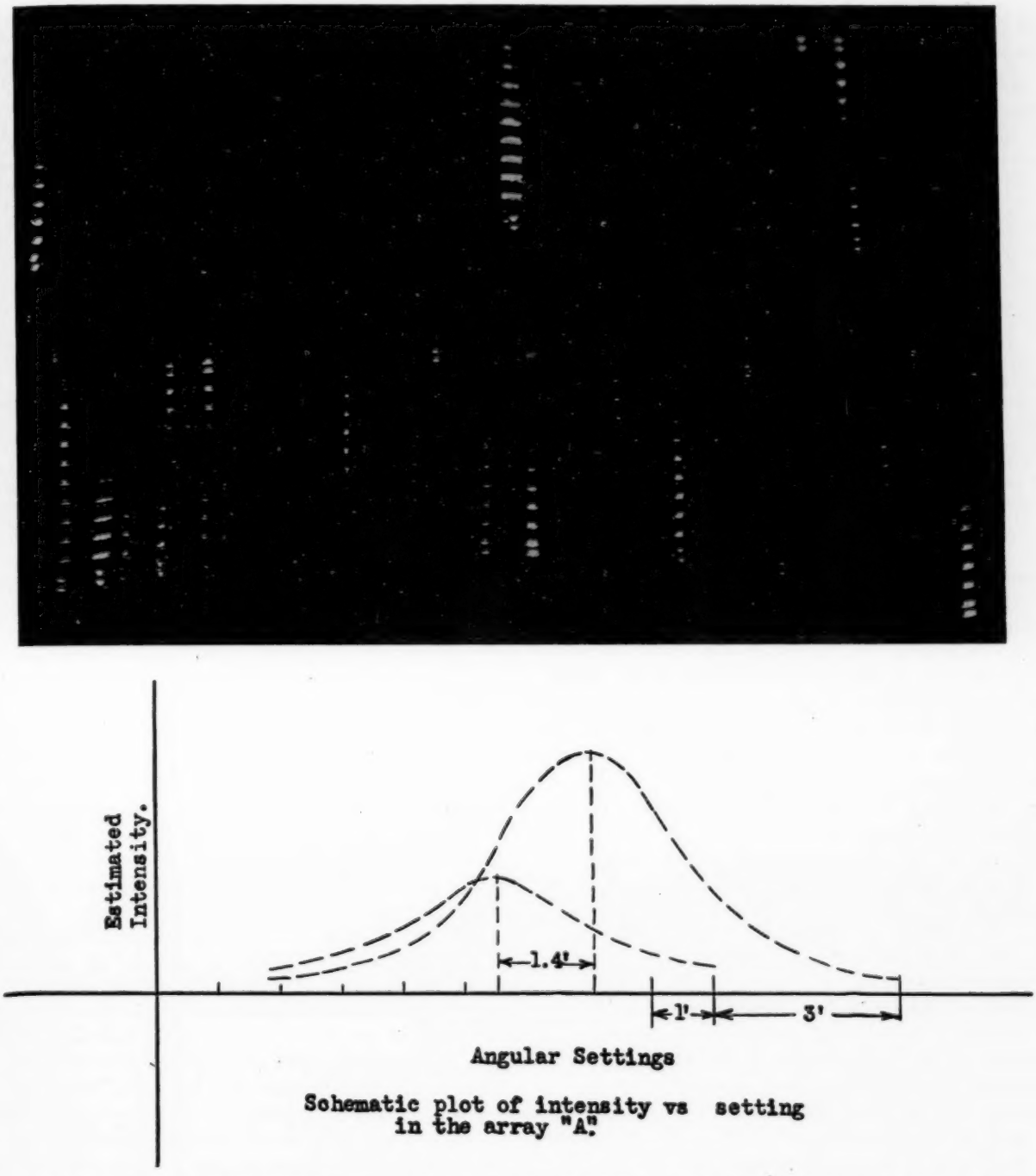


FIG. 4. Experiment 49-M. Multiple exposure of silicon powder. Detailed view of figure 13. Enlargement 8 times. Minus wing. The spectrum includes the (220) and (311) lines. "A," (marked by an arrow) is a typical array of spots. The doublets are well resolved. For each component, the intensity falls off symmetrically with the departure from peak setting. The peaks for both components are approximately 1.4′ apart.

array, a technique was adopted which tacks the terminal exposures by repeating them with a smaller filmshift. This procedure automatically identifies the terminal changes of setting.

As has been pointed out before, the intensities in an array of spots vary with the setting. For most specimens, the intensity values, when plotted *vs* angular setting, lie on a distribution curve. Different arrays of spots differ greatly in intensity and hence in apparent angular extent of reflection. If, however, the half-widths of the intensity distribution function are measured, one is able to evaluate comparable statistical data.

The technique of multiple exposure is illustrated by two multiple exposure diagrams (Figs. 2 and 4) of a

TABLE I. Observed intensities *vs* setting of specimen. Experiment 49-M. Plus wing, reflection (220).[a,b]

Change in setting		3′		1′		1′		1′		1′		1′		1′		1′		1′		3′	
Exposure No.		1		2		3		4		5		6		7		8		9		10		11	
Ψ		α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
13°09′																16		32		48		8	
12°15′										4	3½	8	4	16	4	100	4	144	5	256	6	400	2
8°52′				18		10		6															
7°04′								4		16		24		32		20		5	2	3	3		10
6°37′		36			36		52		40		24		8										
6°09′		7		80		100		64		12		8	8		9		36		48		40		5
5°44′						7		8		10		16		10		8		7		6	2		
5°28′										3		10		15		16		8		3			
3°04′		80	2	16	8		16		32		36		28		9		4						
0°												1		3		10		16		9		7	
−57′								2		8		36		80		176		128		64			28
−4°43′		5	96		36		20		8		5												
−9°58′		24		16		28		40		30			9		3								

[a] The value 100 has been assigned to a certain intensity, and corresponding values have resulted for the other intensities.
[b] Notation for Tables I–IV. ψ the angle subtended by the normal of the reflecting crystal plane with the equator plane. H' apparent half-width in minutes $H'\cos\psi$ half-width corrected for the velocity factor. Q correction term for the convergence of the irradiating x-rays. H half-width fully corrected. $\Delta H = \bar{H} - H$ deviation of an individual half-width figure from average. $\bar{H}$ average half-width. $\langle \Delta H \rangle$ average deviation of half-width. K vertical convergence of irradiating beam = 15 minutes. L horizontal convergence of the irradiating beam = 0.5 minutes. $\sin\psi = Y/R/\{2\sin\theta[1+(Y/R)^2]^{\frac{1}{2}}\}$. $Q = 0.67\{[K\sin\psi/(\cos^2\psi - \sin^2\theta)^{\frac{1}{2}}] + L\}\cos\psi$. $H = H'\cos\psi - Q$.

commercial silicon powder having crystallite sizes from 5 to 25 microns. The arrays of spots are clearly resolved; the small filmshifts give a visual indication of the incomplete arrays.

Experiment 47-M may be considered as an exploratory one. A critical survey of the diagram (Fig. 2) reveals that the total range of 9 minutes was insufficient to cover fully the range of reflection of most crystallites; the settings were, however, too wide apart to ascertain the variation of the intensity as a function of specimen settings as accurately as desirable.

In the multiple exposure 49-M, more settings were used with smaller changes between settings. In the diagram (Figs. 3 and 4), the variation of the intensities as a function of the setting is evident in full detail, and the entire angular range of reflection is covered for many crystallites.

The filmshifts are made visible in the diagram (Fig. 3) as sharp vertical edges of an irradiated area near the primary beam. Any departure from the specified filmshifts would show up in the set of shadow edges.

Some spots in the diagrams are resolved doublets of α_1 and α_2; their occurrence proves the presence of crystallites with an angular range of reflection covering both the positions required for the components.

Transition between the Components

In those spectral lines where doublet spots are rare, one frequently finds 2 spots in a distance which is the standard distance plus or minus the split of the doublet. For one wing, the sign is plus throughout; for the other one, minus throughout. These pairs derive from different components reflected from the same crystallite in different settings.

In taking multiple exposure diagrams, a technique was employed in which 2 or more x-ray films are superimposed. This technique proved to be doubly advantageous. First of all, the absorption in any one film causes the monochromatic radiation to decrease by a specific factor, which is 3.8:1 for copper K-radiation at normal incidence and the Kodak No-screeen x-ray film (code no. 5133). Thus, the same spot printed on subsequent films provides a calibration for comparing estimated intensities as the absorption factor is known. Secondly, each spot on the Debye-Scherrer curve can be examined on that film where its intensity and background is most suitable, a condition which cannot easily be surmized in advance and which varies within the same diagram.

For the qualitative observation, identification and measurement of the diffraction phenomena which become manifest in a multiple exposure, the technique of enlarged projection turned out to be very successful. A projector with a special filmholder was adapted for this purpose. The 3 superimposed films of the multiple exposure diagrams are cut into strips, and the identical spectral regions of the films are projected side by side onto a screen some 17 feet distant. A 19-fold magnification is obtained, and the spots can be identified by means of dividers adjusted to the various distances of filmshift, which can be taken from the sharp vertical edges of the irradiated area near the trap (see Fig. 3). The width and length of the spots of each array is measured and recorded, and intensity values are assigned to each spot through the aid of the simultaneously projected superimposed film.

As a result of this procedure it is possible to plot the intensity values *vs* angular settings within an array of spots, and to obtain a distribution curve. From this curve, the half-width at half-maximum is computed

TABLE II. Half-width of angular range in experiment 49-M. Plus wing, reflection (220).

ψ	H'	$H'\cos\psi$	Q	H	ΔH	$\%H$	$\bar{H}$	$\%\langle\Delta H\rangle$
+8°52′	3	3	2.1	0.9	0.4	31		
+7°4′	2.5	2.5	1.77	0.7	0.6	46		
+6°37′	4	4	1.7	2.3	1.0	77		
+6°9′	3	3	1.5	1.5	0.2	15.4		
+5°44′	3	3	1.62	1.4	0.1	7.7		
+5°28′	2.75	2.75	1.50	1.25	0.05	3.8		
							1.3	23.7
+3°4′	2.5	2.5	1.05	1.45	0.15	11.5		
0°00′	2	2	0.5	1.50	0.2	15.4		
−57′	2	2	1.2	0.80	0.5	38		
−4°43′	3	3	1.51	1.5	0.2	15.4		
−7°4′	3.2	3.1	1.77	1.3	0.0	0		
−9°58′	3.25	3.2	2.2	1.0	0.3	23		

TABLE III. Half-width of angular range in experiment 49-M. Plus wing, reflection (311).

ψ	H'	$H'\cos\psi$	Q	H	ΔH	$\%H$	$\bar{H}$	$\%\langle\Delta H\rangle$
+7°15′	5′	5′	1.85	3.15	1.58	100.0		
+5°50′	3.5	3.5′	1.57	1.93	0.36	23.0		
+4°46′	3.5	3.5′	1.50	2.00	0.45	28.6		
							1.57′	37.5
+2°48′	2.00′	2.0′	1.00	1.00	0.57	36.3		
+2°24′	2.00′	2.0′	0.95	1.05	0.52	33.1		
+1°36′	2.25′	2.25′	0.82	1.43	0.14	8.9		
−2°1′	1.75′	1.75′	0.80	0.95	0.62	39.5		
−3°3–1	2.25′	2.25′	1.17	1.08	0.49	31.2		

using the mathematical expressions which have been described previously and taking into account all necessary correction items for spots off the equator.

EXPERIMENTAL DATA AND STATISTICAL ANALYSIS

The intensity data necessary for the computation of the half-width of the angular range of reflection are obtained from a multiple exposure diagram as depicted in Figs. 3 and 4.

A typical data sheet exhibiting data from observation of the (220) reflection in a silicon spectrum (plus wing) is reproduced in Table I. In the evaluation of the intensity distribution function of an array of spots, advantage is taken from the fact that, if the peak intensity of one component is observed, the position of the peak intensity of the other component can be computed, since the doublet split of any specific (*hkl*)-reflection is known.

The Tables II–V illustrate the computation of the average half-width. In Tables II and III, the average is based on all the values up to the practical azimuthal limit.

For most (*hkl*)-reflections, the average half-width is virtually the same, about 1.5 minutes, and the average deviation is fairly small. The findings differ, however, for the (111)- and the corresponding higher order reflections. All the crystallites which have produced (444)-reflections in either wing (Table V represents the plus wing) show excessive misalignment. Excessive misalignment occurs frequently in the (333)- and (511)-line; in the (111)-reflection, there is one case among 6 exhibiting a very high width value. The excessive misalignments occurring in several orders of reflection of the (111)-plane are too widely scattered to compute a significant average figure.

TABLE IV. Half-width of angular range in experiment 49-M. Plus wing, reflection (111).

ψ	H'	$H\cos\psi$	Q	H	ΔH	$\%H$	$\bar{H}$	$\%\langle\Delta H\rangle$
+6°17′	3.5	3.5	2.1	1.4	0.05	3.5		
+5°31′	3.0	3.0	1.85	1.15	0.3	20.7		
							1.45	14.5
0°00′	6	6	0.5	5.5[a]				
−4°44′	3.0	3.0	1.7	1.3	0.15	10.3		
−5°31′	3.5	3.5	1.7	1.8	0.35	24.1		
−9°30′	4.5	4.5	2.9	1.6	0.2	13.8		

[a] This figure has been omitted from the calculation of the average characteristics.

TABLE V. Half-width of angular range in experiment 49-M. Plus wing, reflection (444).

		H'	Q	H
+	5°17′	9	5.5	3.5
+	4°29′			[a]
+	2°40′			[a]
+	1°9′			[a]
(+)	23′	5.5	0.9	4.6
(−)	0°00′	11	0.5	10.5
(−)	3°40′	9	1.2	7.8

[a] The angular range of reflection of the crystallite overlaps the range of specimen settings. The intensity distribution departs from the usual type and is characterized by a broad flat maximum.

Broadening effects can be studied by comparing the widths of the spindle-shaped spots. In the diagrams 47-M and 49-M, significant broadening effects occur only at large Bragg angles as expected. To obtain quantitative data, it seems necessary to study the intensity distribution across the spot which varies from spot to spot even for the same (*hkl*)-reflection. These phenomena are now under investigation in both single and multiple exposures.

Silicon powder proved suitable for establishing the technique of multiple exposure, as the entire spectrum can be conveniently examined in both wings. This method, however, is not only applicable to the study of powdered material but also to that of compact polycrystalline bodies, such as sheets, plates, or blocks.

CONCLUSIONS

1. The new double spectrometer method appears to be helpful in the study of small imperfections in polycrystalline materials. It has been demonstrated that the data of the x-ray intensity distribution of a reflecting crystallite as a function of crystallite setting can be utilized for the statistical computation of the angular range of reflection of the crystallite. The angular range of reflection is expressed in terms of half-width values.

The experiments have also confirmed that the method is sensitive enough to record slight variations of the angular misalignment for various (*hkl*)-reflections.

2. In the case of the silicon powder spectrum, the statistical analysis of the experimental data has led to a significant conclusion. For most Debye-Scherrer lines, the average half-width is found to be virtually the same, and the deviations from the half-width value are statistically distributed. Significant departures from the average half-width value of 1.5 minutes occur only in the reflection (111), (333 and 511), and (444), where some or all misalignments are excessive. The conclusion may be drawn that the (111)-plane has a unique function in the silicon structure as far as either the mechanism of growth or cohesion is concerned. This is in good agreement with existing evidence on the crystallography and cleavage behavior of diamond, which has the same type of crystal structure.

Broadening effects increase with the Bragg angle as expected. No simple correlation has been observed between broadening effects and angular range of reflection.

ACKNOWLEDGMENT

The double spectrometer was designed by Dr. David Harker, constructed in the research laboratories of the General Electric Company in Schenectady, and donated to one of the authors (A.J.R.) by the General Electric Company. The authors wish to express their gratitude to Dr. Harker and to Dr. William Coolidge, former vice-president in charge of research.

This investigation has been sponsored by the U. S. Office of Naval Research.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 5 MAY, 1951

Attenuation of Radio Signals Caused by Scattering*

A. H. LAGRONE, W. H. BENSON, JR., AND A. W. STRAITON
The University of Texas, Austin, Texas

(Received November 2, 1950)

An equation is developed for determining the total power scattered from an incident radio beam during its passage through a unit scattering volume. Curves are plotted from which the attenuation of the incident signal caused by scattering may be obtained.

I. INTRODUCTION

RADIO waves being propagated through a medium in which the permittivity varies with distance about a median value are subject to attenuation in the form of energy loss caused by scattering. An equation for the magnitude of the power scattered in a given direction has been developed by Booker and Gordon.[1] This paper takes the Booker-Gordon equation and performs a surface integration to find the total power scattered by a unit volume relative to the power incident on the volume.

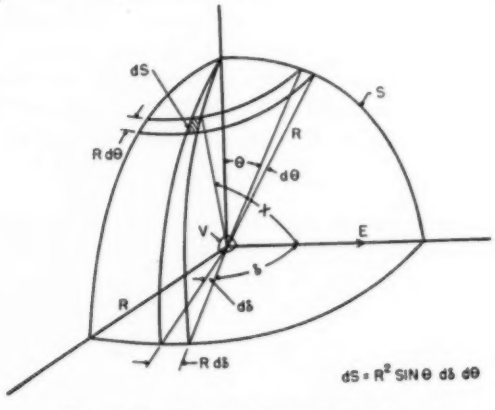


FIG. 1. Geometry of the surface integral.

II. THE BOOKER-GORDON EQUATION

Booker and Gordon deduce the scattered power measured, (i) per unit solid angle, (ii) per unit incident power density, (iii) per unit macroscopic element of volume to be

$$\sigma(\theta, x)=\frac{\langle(\Delta\epsilon/\epsilon)^2\rangle(2\pi l/\lambda)^3\sin^2 x}{\lambda[1+\{(4\pi l/\lambda)\sin\tfrac{1}{2}\theta\}^2]^2}, \qquad (1)$$

where: ϵ = the average permittivity; $\Delta\epsilon$ = the departure of the permittivity from its average value; l = the scale of turbulence; λ = the wavelength; x = the angle between the direction of the electric field vector and the direction of scattering; θ = the angle between the direction of incidence and the direction of scattering.

III. TOTAL POWER SCATTERED BY A UNIT VOLUME

To compute the total power scattered by a unit macroscopic volume, V, per unit incident power density, consider the coordinate system shown in Fig. 1, with V centered at the origin. Let S be a spherical surface of radius R with its center at the origin. If R is very large compared to the largest linear dimension of V, the flow of the scattered power across S will be everywhere normal to S. The total power, w, scattered by V is then the

* This work was sponsored by the ONR at The University of Texas under Contract N5ori-136, P. O. I.

[1] H. G. Booker and W. E. Gordon, Proc. Inst. Radio Engrs. 38, No. 4, 401–412 (1950).

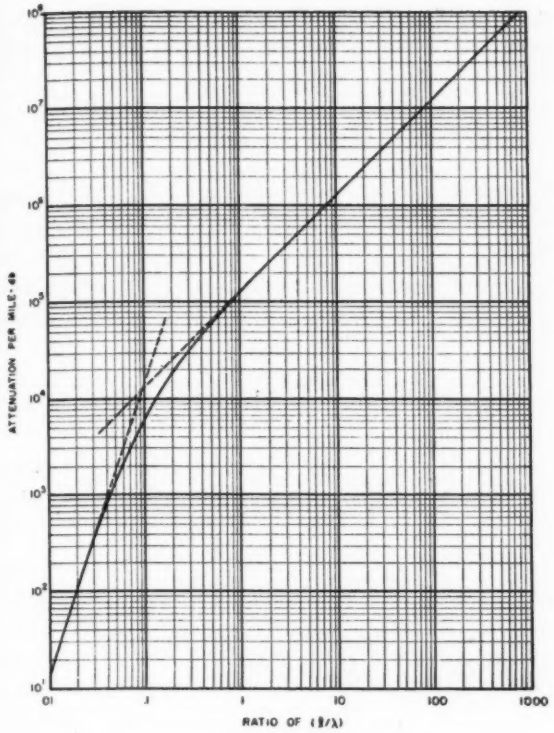


FIG. 2. Attenuation per mile for $[\langle(\Delta\epsilon/\epsilon)^2\rangle/\lambda]=1$.

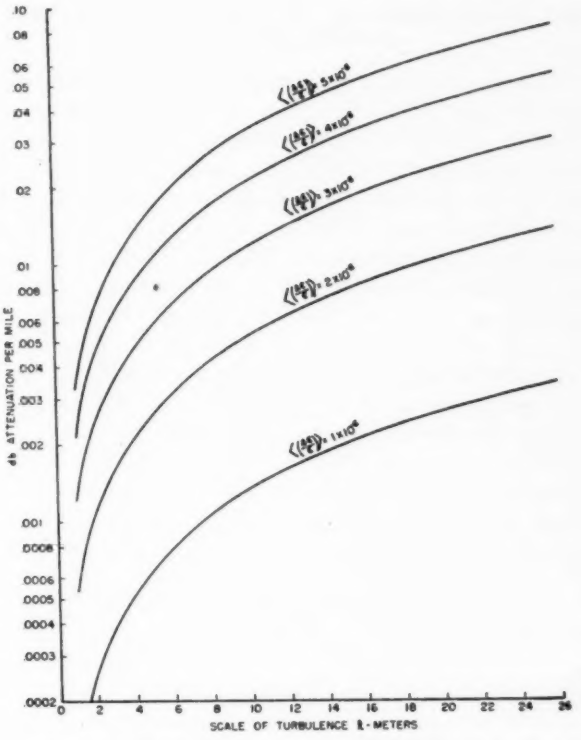


FIG. 3. Attenuation caused by scattering for $\lambda=3.2$ cm.

surface integral of the power density taken over the entire surface S.

The power density, D, at any point (R, θ, x), given by the scattering formula, is

$$D(R, \theta, x)=\sigma(\theta, x)/R^2 \tag{2}$$

and from Fig. 1

$$dS=R^2 \sin\theta d\theta d\delta. \tag{3}$$

Therefore,

$$w=\int\int_S D(R, \theta, x)dS, \tag{4}$$

$$w=\int_{\theta=0}^{\pi}\int_{\delta=0}^{2\pi}\sigma(\theta, x)\sin\theta d\theta d\delta, \tag{5}$$

$$w=\frac{\langle(\Delta\epsilon/\epsilon)^2\rangle(2\pi l/\lambda)^3}{\lambda}\times\int_{\theta=0}^{\pi}\int_{\delta=0}^{2\pi}\frac{\sin^2 x\sin\theta d\theta d\delta}{[1+\{(4\pi l/\lambda)\sin\tfrac{1}{2}\theta\}^2]^2}. \tag{6}$$

From Fig. 1

$$\sin^2 x=\sin^2\delta+\cos^2\delta\cos^2\theta. \tag{7}$$

Substituting for $\sin^2 x$,

$$w=\frac{\langle(\Delta\epsilon/\epsilon)^2\rangle(2\pi l/\lambda)^3}{\lambda}\times\int_{\theta=0}^{\pi}\int_{\delta=0}^{2\pi}\frac{(\sin^2\delta+\cos^2\delta\cos^2\theta)\sin\theta d\theta d\delta}{[1+\{(4\pi l/\lambda)\sin\tfrac{1}{2}\theta\}^2]^2}. \tag{8}$$

Equation (8) integrates to give

$$w=\left(\frac{\pi}{\lambda}\right)\left\langle\left(\frac{\Delta\epsilon}{\epsilon}\right)^2\right\rangle\left[\frac{4(2\pi l/\lambda)^3}{1+(4\pi l/\lambda)^2}+\frac{1}{(2\pi l/\lambda)}-\left\{\frac{2+(4\pi l/\lambda)^2}{8(2\pi l/\lambda)^3}\right\}\log_e\{1+(4\pi l/\lambda)^2\}\right]. \tag{9}$$

This is the total power scattered by the unit macroscopic volume V per unit incident power density.

The attenuation of the incident power caused by scattering can be expressed as follows,

$$\text{db/unit length}=4.34\log_e(P_{\text{in}}/P_{\text{out}}) \tag{10}$$

and for

$$P_{\text{in}}=1,\quad P_{\text{out}}=(1-w),\quad w\langle\langle 1,$$
$$\text{db/unit length}=4.34\log_e[1/(1-w)]\cong 4.34w. \tag{11}$$

If the wavelength is measured in meters the attenuation per mile will be

$$\text{db/mile}=(4.34)(1609)w=6983w, \tag{12}$$

$$\text{db/mile}=6983\frac{\pi}{\lambda}\left\langle\left(\frac{\Delta\epsilon}{\epsilon}\right)^2\right\rangle\left[\frac{4(2\pi l/\lambda)^3}{1+(4\pi l/\lambda)^2}+\frac{1}{(2\pi l/\lambda)}-\frac{2+(4\pi l/\lambda)^2}{8(2\pi l/\lambda)^3}\log_e\{1+(4\pi l/\lambda)^2\}\right]. \tag{13}$$

IV. CHARACTERISTICS OF THE TOTAL SCATTERED POWER EQUATION

The curve in Fig. 2 shows a plot of Eq. (13) for a very special case of the variables. Since it would not be feasible to show a complete family of curves for all possible combinations of the variables, Eq. (13) is plotted for a rather large range of the variable (l/λ) with $\langle(\Delta\epsilon/\epsilon)^2\rangle$ and $(1/\lambda)$ set equal to unity. The db attenuation per mile for any case can then be found by simply taking the value of Eq. (13) from the curve and multiplying it by the measured or assumed values of $\langle(\Delta\epsilon/\epsilon)^2\rangle$ and $(1/\lambda)$. The unit for λ in this case is meters.

It is interesting to note that the curve in Fig. 2 could be very closely approximated by two straight lines on the log log coordinate paper. The reason for this can be found by examining Eqs. (8) and (13).

For (l/λ) large (in the order of 1) the second and third terms in Eq. (13) are small compared to the first term and the db attenuation per mile is approximately

$$\text{db/mile} \cong 6983\left(\frac{\pi}{\lambda}\right)\left\langle\left(\frac{\Delta\epsilon}{\epsilon}\right)^2\right\rangle\left[\frac{4(2\pi l/\lambda)^3}{(4\pi l/\lambda)^2}\right], \quad (14)$$

$$\text{db/mile} \cong 6983(\pi/\lambda)\langle(\Delta\epsilon/\epsilon)^2\rangle(2\pi l/\lambda), \quad \text{for} \quad (l/\lambda) \geqq 1. \quad (15)$$

For (l/λ) small (in the order of .01) Eq. (8) is approximately

$$w \cong \frac{(2\pi l/\lambda)^3}{\lambda}\left\langle\left(\frac{\Delta\epsilon}{\epsilon}\right)^2\right\rangle \times \int_{\delta=0}^{\pi}\int_{\delta=0}^{2\pi}(\sin^2\delta + \cos^2\delta\cos^2\theta)\sin\theta d\theta d\delta. \quad (16)$$

After integrating Eq. (16) and multiplying by 6983 to convert to db/mile, we have

$$\text{db/mile} = \frac{6983}{\lambda}\left(\frac{8\pi}{3}\right)\left\langle\left(\frac{\Delta\epsilon}{\epsilon}\right)^2\right\rangle(2\pi l/\lambda)^3, \quad \text{for} \quad (l/\lambda) \leqq .01. \quad (17)$$

Equation (15) shows the linear relationship found in the top portion of the curve while Eq. (17) shows the cubic relationship of the lower portion.

V. APPLICATION TO SPECIFIC CASES

As an example of the attenuation due to scattering, a wavelength of 3.2 cm was chosen and several curves computed for various values of $\langle(\Delta\epsilon/\epsilon)\rangle$. These curves are shown in Fig. 3. The range covered by the variables $\langle(\Delta\epsilon/\epsilon)\rangle$ and (l) is chosen to agree with values currently reported.[1–3]

From Fig. 3 it is apparent that attenuation caused by scattering at $\lambda = 3.2$ cm is negligible over the range of the variables considered, and for $\lambda > 3.2$ cm, Eq (13) shows that the attenuation would be smaller than that shown. It follows then, for the values of average index change and scale of turbulence considered by Booker and Gordon,[1] that the attenuation of the incident beam caused by scattering is negligible. This justifies their assumption that multiple scattering can be neglected. It should be recognized, however, that little information is available on the magnitude of the scale of turbulence and magnitude of index changes found throughout the troposphere.

VI. SUMMARY

An equation has been obtained which gives the total power scattered per unit volume per unit incident power density.

The power scattered per unit volume is found to vary directly as the scale of turbulence and as the square of the frequency for (l/λ) large, and as the scale of turbulence cubed and as the frequency to the fourth power for (l/λ) small.

---

[2] F. J. Scrase, "Some characteristics of Eddy motion in the atmosphere," Geophysics Memoir of London Meteorological Office, No. 52 (1930).

[3] J. R. Gerhardt and W. E. Gordon, J. Meteorology **5**, 192–203 (1948).

On the Diffraction of Electromagnetic Waves by Two Conducting Parallel Half-Planes*

M. G. CHENEY, JR.,† AND ROBERT B. WATSON
Department of Physics and Defense Research Laboratory, The University of Texas, Austin, Texas
(Received January 19, 1951)

An examination is made of the diffraction of radar waves produced by the edges of two parallel half-planes, disposed one behind the other relative to the signal source. A number of experimental patterns are exhibited. Computed patterns are obtained following a line of analysis formulated by Baker and Copson. Comparison between experiment and theory shows only fair agreement, indicating the need for a better theoretical approach.

INTRODUCTION

THE use of radar waves for obtaining diffraction patterns has been demonstrated in a number of papers.[1] Earlier work has been largely concerned with single or coplanar obstacles. However, diffraction by noncoplanar obstacles becomes of considerable importance for antennas mounted near metallic surfaces on aircraft and elsewhere. There have been several recent investigations of optical diffraction from noncoplanar surfaces using conventional optical theories for interpretation of experimental data.[2] In addition, several articles of a theoretical nature have appeared recently which consider the diffraction of electromagnetic waves by noncoplanar perfectly conducting planes.[3] The present work considers the diffraction produced by the edges of two semi-infinite perfectly conducting planes mounted perpendicular to the line of propagation of plane parallel electromagnetic waves. The edges of these planes are taken to be parallel, and one of the planes is behind the other with reference to the oncoming waves but is more or less shaded by the first plane. Diffraction patterns are obtained about the edge of the plane furtherest from the source.

APPROXIMATE ANALYSIS

The diffraction field about the edge of a single semi-infinite perfectly conducting plane has been analyzed for a plane electromagnetic wave by Baker and Copson.[4] The resultant field is considered to be made up of combinations of the direct and reflected waves, as found by geometrical optics, and a cylindrical wave with its origin at the edge of the semi-infinite plane. For analysis, the region about the plane is divided into three parts: the geometrical shadow region, the region between the source and the plane (region of geometrical reflection), and the illuminated region in which the direct wave passes the edge. The analysis leads to rather simple results in these regions, except near the boundaries (other than at the plane itself) where more extended analysis is required for parabolic subregions having the edge as foci and the geometrical boundaries of the major regions as principal axes.

For the present problem, this analysis has been adopted but with the consideration of additional regions. These regions are shown in Fig. 1. In Region I, it is assumed that the only wave of importance is the cylindrical wave originating at the edge of Plate 2. This diffracted wave is computed as though Plate 1 were absent. In Region II, there are two waves: the direct wave from the source and the cylindrical wave from

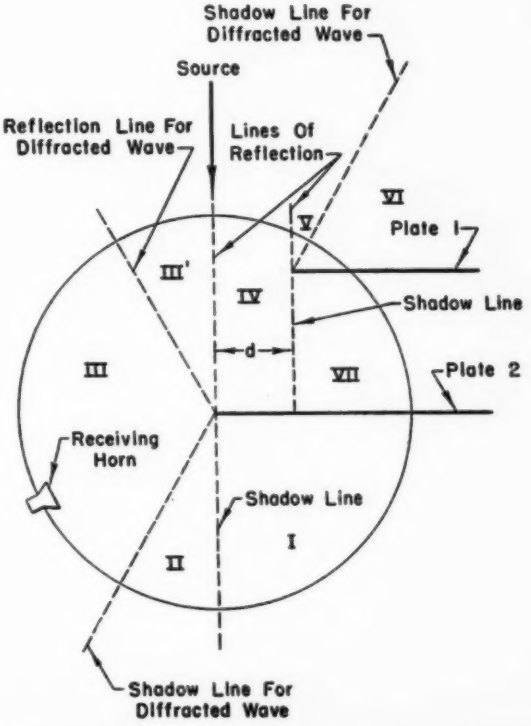


FIG. 1. Regions about diffracting edges for analysis.

* The material in this paper is drawn largely from the thesis, "Diffraction of electromagnetic waves by two conducting parallel half planes," presented by M. G. Cheney, Jr., in June, 1950, in partial fulfillment of the requirements for the MA degree. The experimental work and the computations were carried out at the Defense Research Laboratory under the sponsorship of the Bureau of Ordnance, Navy Department, Contract NOrd-9195.

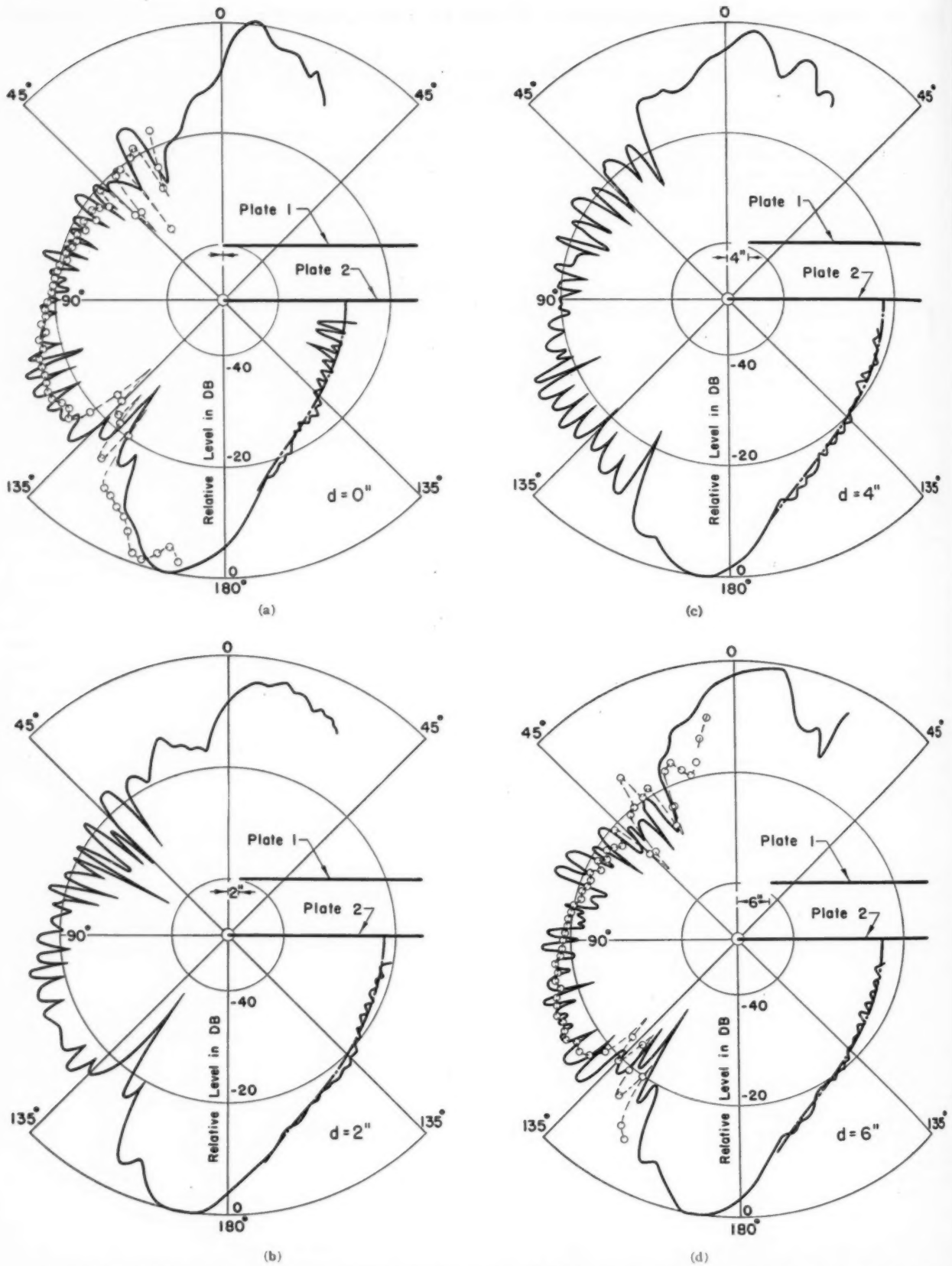
† Now at Columbia University.

[1] For a similar consideration to that presented in this paper, but for a semi-infinite plane, and for a semi-infinite wedge, see C. W. Horton and R. B. Watson, J. Appl. Phys. **21**, 16 (1950); R. B. Watson and C. W. Horton, J. Appl. Phys. **21**, 802 (1950).

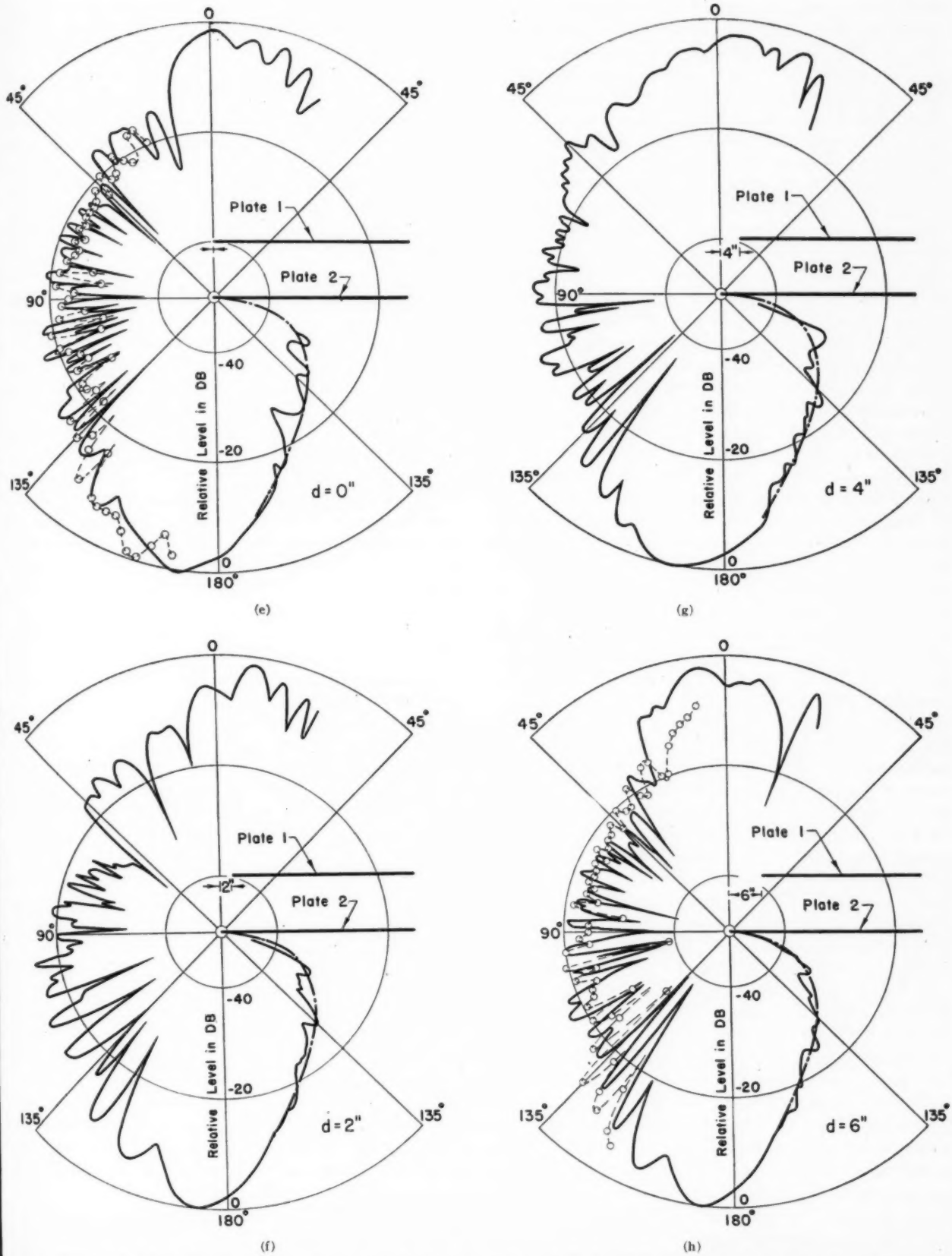
[2] Carlos F. Ellis, Am. J. Phys. **16**, 8 (1948); James R. Heirtzler, Am. J. Phys. **17**, 419 (1949).

[3] F. E. Carlson and A. E. Heins, Quart. Appl. Math. **4**, 313 (1947); **5**, 82 (1947); A. E. Heins, Quart. Appl. Math. **6**, 157 and 215 (1948); **8**, 281 (1950).

[4] B. B. Baker and E. T. Copson, *Mathematical Theory of Huygens' Principle* (Oxford University Press, London, 1939), chapter 4.



FIGS. 2(a)–(d). Diffraction patterns for two conducting noncoplanar half-planes. *E*-vector perpendicular to diffracting edges; parameter, plate separation d; solid curve, experiment; dashed curves and circles, theory.



FIGS. 2(e)–(h). Diffraction patterns for two conducting, noncoplanar half-planes. *E*-vector parallel to diffracting edges; parameter, plate separation d; solid curve, experiment; dashed curves and circles, theory.

Plate 2. In addition to these two waves, there must be included in Region III the cylindrical wave arising from Plate 1. This cylindrical wave is found from the total field at the edge of Plate 1, assuming that this field is made up from the direct wave from the source and the first reflected wave from Plate 2. For the calculation of this first-reflected wave, Plate 2 is considered to act as an infinite plane. In Region III′, there appears, further, the cylindrical wave from Plate 1 as reflected in Plate 2. In Region IV, there is yet another additional wave, the direct wave reflected from the exposed portion of Plate 2. In Region V, the reflected wave of Plate 2 is replaced by the reflected wave of Plate 1. In Region VI, the cylindrical wave of Plate 2 is also absent. Finally, in Region VII, only the cylindrical waves are present. As in the analysis of the diffraction about the edge of a single semi-infinite plane, moderately simple results are obtained except for parabolic subregions at the boundaries of these major regions where more extended analysis is required.

In summary, the simplifying assumptions in the analysis are given as follows. The plates can be considered as perfectly conducting, of zero thickness, and of semi-infinite extent. The Plates are considered perfectly flat, with the active edges parallel. The Plate 2 is considered to be infinite in extent only in computing diffraction produced by the edge of Plate 1 from the reflected wave from Plate 2. The wave guide effect of the parallel planes is considered negligible, except for the inclusion of the first reflection of the direct wave from Plate 2 and the first reflection of the diffracted wave from Plate 1 at Plate 2.

The analysis therefore uses the results of Baker and Copson directly, care being taken in the addition of several waves to observe appropriate phase angles because of the physical separation of the two edges. In the experimental work, the receiver was rotated about the edge of Plate 2 as an axis. Accordingly, the diffracted wave from the edge of Plate 2 was received with full strength, while all other waves were affected by the directionality of the receiver in accordance with the angle from the line of maximum reception of the receiver (towards the edge of Plate 2) to the line connecting the receiver and the effective source in each case. In the analysis, appropriate factors due to the receiver's directional pattern were applied to the expressions for the various waves.

Detailed calculations from the resulting formulas were made for the Regions I and III, except for the parabolic subregions, for two relative positions of the two plates. Figure 2 includes plots of these computed values which were made at intervals of 2.5°. Intermediate points can be estimated reasonably well by an examination of the amplitudes of the several combining waves, ignoring the phase factors. Hence, the plots show both the computed points and an estimate of the curve for the intermediate points. Calculations for the other regions and relative positions were not carried out because of the considerable labor of the computation, in view of the moderate agreement between experiment and theory.

EXPERIMENT

The general experimental arrangement has already been described in an earlier paper.[1] The two semi-infinite plates were made of two sheets of copper, each 30 in.×60 in.×$\frac{1}{16}$ in. thick, supported as earlier described except for Dural channel members placed between the support members for the two sheets to produce the desired separation of $12\frac{3}{4}$ in. (about ten wavelengths). The plate which was nearest to the source was arranged so that it could be slid along its length for a few inches in order to vary the distance of separation (d in Fig. 1) of the two edges. The plate furtherest from the source was fixed in position, and the receiving horn was rotated about the edge of this plate on a circle of $15\frac{1}{2}$ in. radius.

Experimental data were obtained at times when the wind was absent, or nearly so, because of the motion of the plates caused by even a slight breeze. Many of the observations were, for this reason, taken at night; some effect was observed because of the formation of dew on the plates, but on most occasions the plates were dry. Regrettably, there were others with a vital interest in dish reflectors and wave guides. On a quiet evening, apparent wind effect led to an investigation of the transmitter, and it was discovered that two spiders were busily stringing a web from antenna to dish rim. Cleaning out spiders and webs was, thereafter, a regular procedure.

Diffraction patterns were obtained for both vertical and horizontal polarization; that is, with the electric vector both parallel to and perpendicular to the edges of the plates, for values of the separation (d) of the edges of the plates of 0, 2, 4, and 6 in. The eight resulting patterns are shown in Fig. 2.

CONCLUSIONS

In all of the experimental curves shown in Fig. 2, the general shape is set by the directional pattern of the receiving horn, in discrimination against the incident plane wave. The effects of the various diffracted and reflected waves are apparent in the variations from a smooth curve, together with a shifting of the position of the major lobe (near 180°) into the region of illumination from the source, and the reduction of intensity in the shadow. On the basis of the assumptions made, the field in Region I should be the same for all relative positions of the two plates. Therefore, the same computed curve is fitted to the experimental data given in Figs. 2a–d which are for the electric vector at right angles to the diffracting edges. As might be expected, the deviations from the theory are greatest for no separation (Fig. 2a) where there is the best chance for the diffraction from Plate 1 to appear in Region I. However, some variations in the data are observed for other

separations, indicating the presence of incidental diffraction from the other edges of the plates and from the support members. A visual fit between theory and experiment is made throughout in the absence of a normalizing factor which would be obtained in the neighborhood of the maximum; the more rigorous calculations were not made in this critical region, as was pointed out earlier. The same order of deviations is observed for the electric vector parallel to the diffracting edges, as shown in Figs. 2e–h.

The fields in the Region III are functions of the separation d of the two plates. Computed values are shown in Fig. 2a for $d=0$ and in Fig. 2d for $d=6$ inches, for the electric vector perpendicular to the diffracting edges. The agreement is not too good; the computations give smaller variations than are observed in practice. The experimental error is not greater than 2 db, whereas the observed deviations are considerably greater in amount. Better agreement is observed in Figs. 2e and 2h, for $d=0$ and 6 in., respectively, for the electric vector parallel to the diffracting edges. Again, the measured fluctuations are somewhat greater than the computed values.

It is concluded that the theoretical approach using the expressions developed by Baker and Copson is not sufficiently accurate to specify completely the diffraction arising from the edges of two parallel and noncoplanar conducting plates. Satisfactory agreement is observed in the shadow zone, but better agreement might be desired in the illuminated zone. A more complete analysis, perhaps using the general methods suggested by Heins,[3] should be more satisfactory.

Publication Charge for Letters to the Editor

At the meeting of the Publication Board of the American Institute of Physics on January 30, 1951, a proposed publication charge for Letters to the Editor (and similar sections) was approved. The charge will be effective with the July issues and will be at a flat rate of $10.00 for each Letter. As in the case of regular articles, the charge will be made against the institution rather than the individual author and will be on a voluntary basis. When honored, 100 reprints will be provided without further cost.

The Journals affected are the following:

Physical Review
Journal of the Acoustical Society
Review of Scientific Instruments (Laboratory and Shop Notes)
Journal of Chemical Physics
Journal of Applied Physics

Multiple-Beam Interferometry

Professor S. Tolansky of the University of London will deliver a series of lectures on Multiple-Beam Interferometry at the Squier Signal Laboratory, Fort Monmouth, New Jersey, on June 28 and 29, 1951. Members of the American Physical Society will be welcome to these lectures. Those interested in attending should address the Director, Squier Signal Laboratory, for further details before June 9.

Letters to the Editor

Radiofrequency Mass Spectrometer

PAUL SCHISSEL
General Electric Company Engineering Laboratory, Schenectady, New York
(Received February 21, 1951)

SEVERAL types of radiofrequency mass spectrometers have been proposed using either velocity selection of the ions,[1] time of flight selection,[2] or a resonance effect in crossed fields.[3] The following spectrometer would utilize the natural period of ions in a parabolic potential distribution. A. M. Skellett[4] has proposed an electron accelerator which utilizes a similar principle.

As shown in Fig. 1, a dc voltage, V_0, is distributed along the

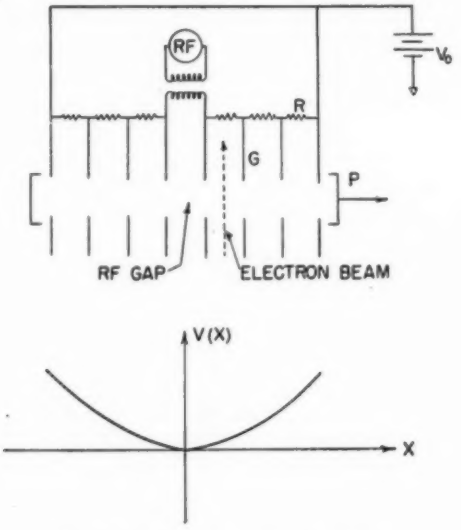


FIG. 1. Radiofrequency mass spectrometer.

axis of the tube by the resistors, R, and rings, G. Ions formed along the electron beam will be pulsed by the rf as they pass through the gap. The ions will move along the axis of the tube until their kinetic energy is converted into potential energy. They will then move in the opposite direction; and if their period of oscillation in the potential well is an odd integer times the period of the rf, they can continue to gain energy from the oscillator.

If the potential distribution is of the form

$$V(x)=\tfrac{1}{2}K^2x^2,$$

the period of oscillation in the potential well is independent of amplitude. Hence, an ion with correct period will continue to gain energy from the rf until it escapes from the potential well and reaches the collector, P.

The equation of motion of an ion is

$$md^2x/dt^2=-e\partial V/\partial x,$$

or

$$x(t)=A\sin[(eK^2/m)\tfrac{1}{2}t-B],$$

where the following definitions hold: m=mass of ion, e=charge on ion, x=distance from center of tube, V=potential distribution along tube, K=a constant.

Hence, the period of the ion motion is

$$T_{\text{ion}}=2\pi(m/eK^2)^{\frac{1}{2}}=(2j-1)T_{rf};$$

and if we solve the above equation for m, we obtain the basic equation for the spectrometer:

$$m=(2j-1)^2eK^2/4\pi^2f^2,$$

where f=frequency of oscillator voltage and j=an integer.

The resolution of the spectrometer may be computed to be

$$\text{Res}=(2j-1)\pi V_0/2E,$$

where V_0=dc supply voltage and E=peak value of the oscillator voltage.

[1] W. H. Bennett, J. Appl. Phys. **21**, 143 (1950).
[2] S. A. Goudsmit, Phys. Rev. **74**, 622 (1948).
[3] Hipple, Sommer, and Thomas, Phys. Rev. **76**, 1877 (1949).
[4] A. M. Skellett, J. Appl. Phys. **19**, 187 (1948).

Oriented Crystals in the Rolled Plates of Alloys

SHIGETO YAMAGUCHI
Scientific Research Institute, Kamifujimae Komagome, Bunkyo-ku, Tokyo, Japan
(Received November 16, 1950)

THE etched surfaces of rolled plate of stainless steels (18-8-Mo and 18-8-Mo-Cu) were detected by means of electron diffraction and oxide replica method of electron microscopy. The etching solution used here was CH_3OH-Br (50:1 by vol). The diffraction patterns (Figs. 1 and 2) obtained from the etched surfaces showed the existence of face-centered cubic austenites, whose (111) planes were oriented in 20°–30° for the surfaces of

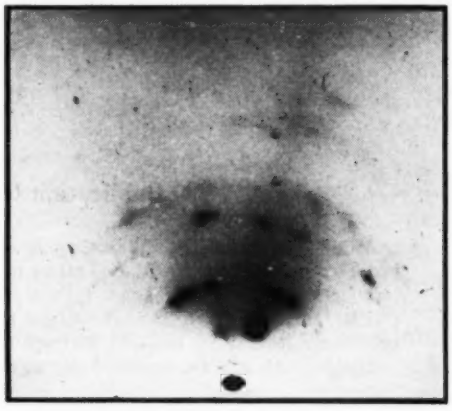


FIG. 1. Diffraction pattern from the etched surface of 18-8-Mo(Mo: 2 percent).

test pieces. The etched surfaces of test pieces were oxidized with water vapor of high pressure (15–20 atmos) in an autoclave. The oxide films stripped from the substrates in a CH_3OH-Br solution were amorphous (structureless) enough to be used as replicas for electron microscopy. The micrographs obtained are shown in Figs. 3 and 4. The boundary planes of a single crystal recognized in Fig. 1 might be assigned with (001) and (111). Since the electron beam in observing Fig. 1 was nearly perpendicular to the replica

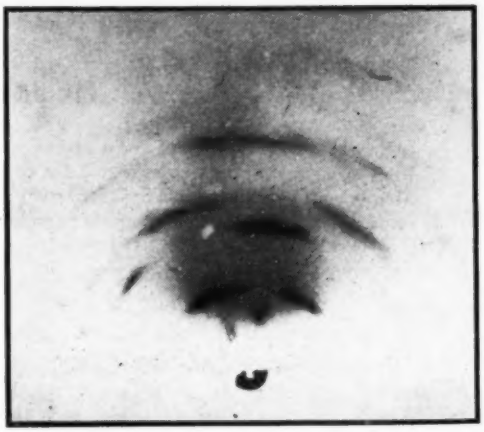


FIG. 2. Diffraction pattern from the etched surface of 18-8-Mo-Cu(Mo, Cu: 2-3 percent).

FIG. 3. Single crystal found on the etched surface of 18-8-Mo.

film, it was possible to determine the angle (30.5°) between the (111) planes and the surface plane of test piece through simple geometric analysis of Fig. 3. This evaluated angle agreed with the result of electron diffraction. There are discernible in Fig. 4 the oblique boundary planes of single crystals. Electron diffraction and microscopy make it possible to determine the indices and orientation of the boundary planes found on the etched surfaces of metals.

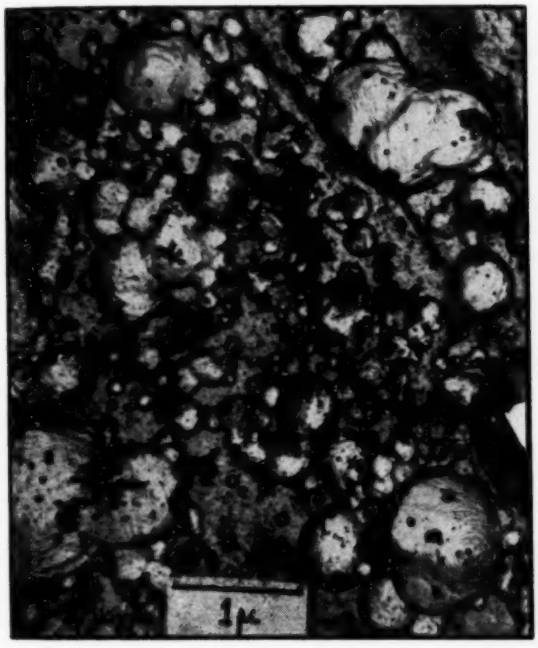


FIG. 4. Etching pits found on the surface of 18-8-Mo-Cu.

Studies of Thin Films by Electron Diffraction

N. R. MUKHERJEE AND OLIVER ROW
Engineering Experimental Station, University of Washington, Seattle, Washington

(Received March 12, 1951)

IN attempts to prepare samples for electron diffraction powder patterns, studies of the difficulties involved led to a method for making samples under controlled conditions. This method provided an opportunity to study the optimum conditions for getting good diffraction patterns.

It is known that the width of the diffracted ring is a function of particle size.[1] We find that samples prepared in the following way show diffraction patterns with narrow (i.e., sharp) rings, a greater number of rings than usual, and a better contrast between the rings and the background at a critical thickness.

Most methods of preparation of diffraction samples use a substrate as supporting medium, e.g., collodion, Formvar, etc. Other methods use cellulose compounds or some crystals, which may be dissolved away leaving the sample in pure form. In the present technique a water soluble plastic film, such as Elvanol* is used as a substrate upon which a film of material is deposited by evaporation as described below. A 15 percent aquous solution of Elvanol is spread evenly on a flat glass plate and allowed to dry slowly. On drying, the film is peeled off, cut into proper dimensions, and stuck to a flat plate (2-cm by 1-cm aluminum plate) using moisture as a cement. Nine such supported Elvanol films are placed tangentially on the arcs of nine concentric circles, the center of which is at the point source of evaporation. A known weight of metal, which is put in a conical filament of tungsten or other suitable material,[2] is located at the horizontal plane passing through the centers of all the nine films mounted vertically. The metal is evaporated completely under high vacuum ($>10^{-4}$ mm Hg). Then each metal film with the substrate is peeled off and laid on the surface of an undisturbed water bath for dissolution of the substrate. The diffraction samples are made from the substrateless floating film by placing an electron microscopic screen disk on a flat-top brass cylinder and bringing it from under the floating film catching a piece of film on the screen. After slow drying, the screen with the sample is placed in the electron diffraction camera.

For illustration of a typical sample, the diffraction pattern of aluminum as obtained by evaporation is shown in Figs. 1 to 3. The results are reproducible. Approximate thicknesses as calculated by inverse square law are 1300A, 580A, and 240A, respectively. The diffraction pattern in Fig. 2 is definitely superior to those in Figs. 1 and 3, with respect to sharpness of the rings, number of rings, and contrast of the rings against the background. Absence of substrate eliminates inelastic scattering caused by it and produces diffraction pictures without undesirable haziness. Most of the published diffraction pictures of aluminum show about eight rings,[3] whereas twenty-four rings are found in Fig. 2 (in the negative there are twenty-seven rings). From the sharpness of the

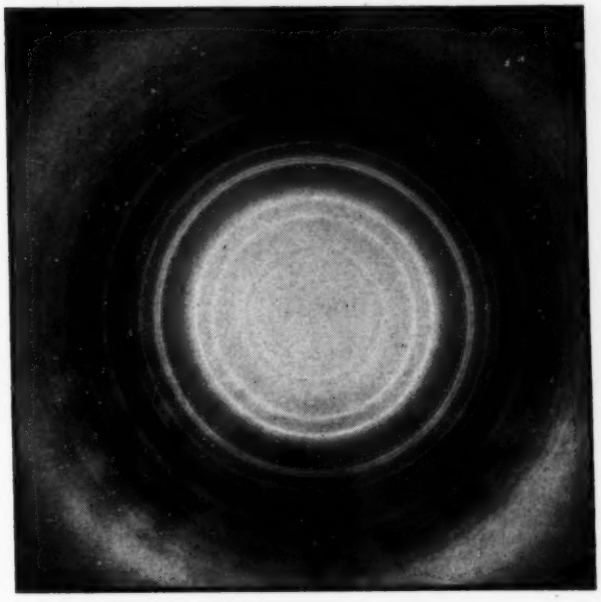


FIG. 1. Diffraction rings of Al—1300A.

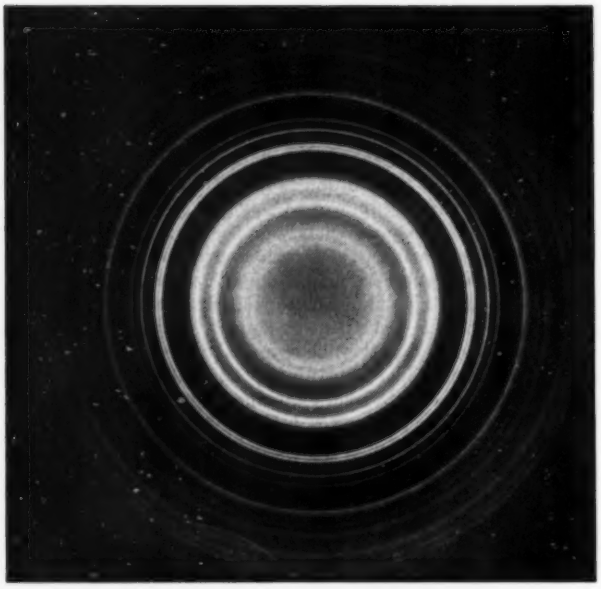


FIG. 2. Diffraction rings of Al—580A.

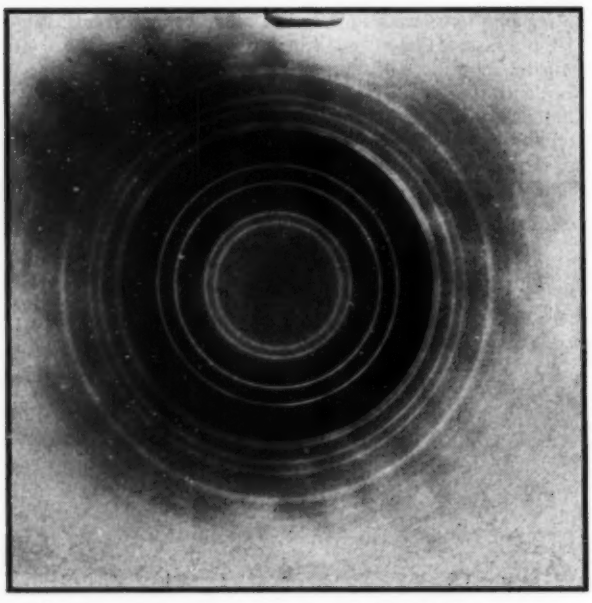


FIG. 3. Diffraction rings of Al—240A.

rings it is believed that the particles obtained by evaporation are of appropriate size.

Further studies will be published later.

[1] G. P. Thomson and W. Cochrane, *Theory and Practice of Electron Diffraction* (Macmillan and Company, Ltd., London, 1939), p. 78.
[*] Elvanol (grade 51-05) was supplied by courtesy of DuPont de Nemours and Company.
[2] W. C. Caldwell, J. Appl. Phys. **12**, 779 (1941).
[3] Simard, Burton, and Barnes, J. Appl. Phys. **16**, 832 (1945).

Comment on Nomenclature in "Electron Microradiography of Electrodeposited Metals"

G. M. CORNEY
Kodak Research Laboratories, Rochester, New York
(Received March 7, 1951)

A RECENT letter[1] from R. Weil and H. J. Read to the *Journal of Applied Physics* concerns a method of preparing metal samples and examining them in the electron microscope. The authors suggest the term "electron microradiography" for their technique. This raises a question of nomenclature, since the phrases "electron (electronic) radiography (microradiography)"[2-7] are already in use to designate a different technique. In this, a thin specimen (e.g., paper) is tightly sandwiched between a sheet of metal foil, usually lead, and a photographic film. The whole is then irradiated by hard x-rays. The electrons emitted from the foil by the action of the x-rays are differentially absorbed by the specimen and record its structure on the film.

In some papers in French[8,9] the terms "(micro) radiographie par electrons secondaires" have also been used to designate this this technique. It seems likely that this phrase would also be rendered into English as "(secondary) electron radiography."

It would appear that the conventional terms "electron microscopy" and "electron micrograph" could be applied to examination of samples prepared by the new technique described by Weil and Read.

[1] R. Weil and H. J. Read, J. Appl. Phys. **21**, 1068 (1950).
[2] H. E. Seemann, J. Appl. Phys. **12**, 836 (1937).
[3] H. S. Tasker and S. W. Towers, Nature **156**, 50 (1945).
[4] J. J. Trillat and C. Legrand, Compt. rend. **224**, 655 (1947).
[5] C. Legrand and J. J. Trillat, Compt. rend. **224**, 1000 (1947).
[6] J. J. Trillat, J. Appl. Phys. **19**, 844 (1948).
[7] D. E. Beischer, Science **112**, 535 (1950).
[8] A. Saulnier and J. J. Trillat, Compt. rend. **220**, 772 (1945).
[9] A. Saulnier and J. J. Trillat, Rev. sci. **83**, No. 3244, 9211 (1945).

Effect of Auxiliary Current on Transistor Operation

H. J. REICH, P. M. SCHULTHEISS, J. G. SKALNIK, T. FLYNN, AND J. E. GIBSON
Yale University, New Haven, Connecticut
(Received March 7, 1951)

IN the course of a Navy-sponsored research project on transistor trigger circuits, the writers have found that under certain conditions the characteristics of transistors are improved by the flow of direct current between auxiliary electrodes, placed and connected as shown in Fig. 1. The most favorable results have

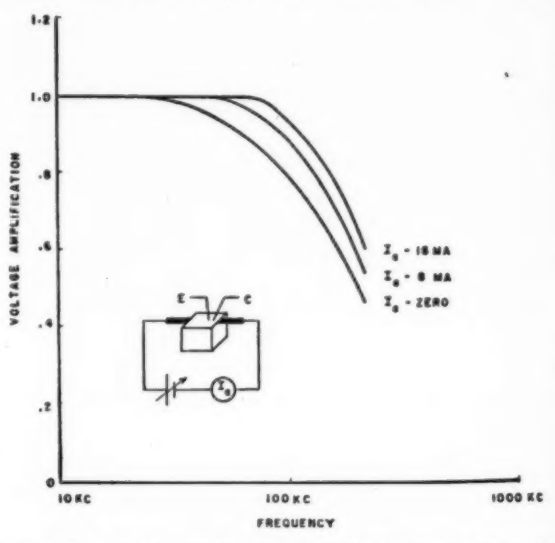


FIG. 1. Effect of the auxiliary current on the frequency response of a transistor amplifier.

been obtained when the auxiliary electrode adjacent to the collector is placed as close as possible to the collector. The auxiliary electrodes are designed to make contact over an appreciable area of the germanium surface.

The curves shown in Fig. 1 were obtained with the transistor used in a grounded-base amplifier circuit. The frequency characteristic of the transistor circuit without bias current was not

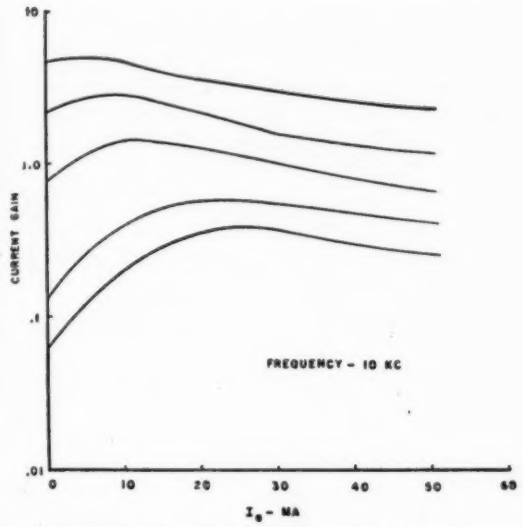


FIG. 2. Improvement in current gain as a function of the auxiliary current. The various curves are for different emitter-collector spacings.

particularly good, but the curves show the improvement if frequency response produced by the auxiliary current. These curves have been normalized to unity at 10 kc.

The manner in which the improvement of current gain depends upon emitter-collector spacing is shown by the curves of Fig. 2. To obtain the data for these curves, the position of the emitter point was changed after formation of the collector contact. The current gain was measured by monitoring the voltage across 100-ohm resistors in the emitter and collector circuits. The curves of voltage gain *vs* auxiliary current are similar to those of current gain. The curves of Fig. 2 indicate that considerable improvement results from the auxiliary current when the emitter-collector spacing is relatively large, but that the improvement is negligible for close spacing. The curves of Fig. 2, and similar curves obtained with other germanium samples, appear to indicate that the improvement resulting from the auxiliary current is, in general, an inverse function of the initial gain.

The position of the probe on the emitter side of the germanium block does not appear to be critical. Although the curves shown in Fig. 2 were obtained with two probes, subsequent experiments suggest that the same results can be obtained by connecting to the base instead of to the probe adjacent to the emitter.

The auxiliary circuit may also be used to achieve an amplitude-modulated output in the collector circuit.

Books

Heat and Temperature Measurement

BY ROBERT L. WEBER. Pp. 422+x, Figs. 14.10 in Part I, 43 in Part II. Prentice-Hall, Inc., New York, 1950. Price $5.00.

In this book the author has presented many valuable concepts and principles which are vital to the measurement of temperature and heat. They are stated in a clear and logical manner. The use of changes in the physical properties of a substance as the basis for temperature measuring devices is discussed and illustrated with many thermometric substances. Chapter 2 is devoted to the use of expansion as a temperature measuring property, and many instruments using both linear and volume expansion are thoroughly explained. A chapter is then devoted to a rather advanced and thorough discussion of the conduction of heat under various conditions. In Chapter 4 the sources of thermal emf are presented and discussed; their thermodynamic relations are given, the relations between emf and temperature, and the thermoelectric power as the rate of change of thermal emf with temperature.

In Chapter 5 the author discusses the use of the above thermoelectric effects in various forms of temperature measuring instruments. The various thermocouple metals, their characteristics and limitations, are fully described. Considerable attention is given to electric circuits used, to junction compensation and the calibration of the thermocouples.

The laws of radiant energy are thoroughly discussed in Chapter 6, which is then followed by a chapter devoted to their use in optical and radiation pyrometry. The next chapter is devoted to resistance thermometry, and is followed by a chapter describing both the principles employed and the forms of many kinds of devices used in temperature recording and control.

In Chapter 10 the author discusses calorimetry and describes a number of types of calorimeters and their uses. This is a particularly valuable chapter to the student who must make some accurate determinations of quantities of heat. This chapter is followed by one devoted to phase diagrams and thermal analysis and another devoted to thermodynamics.

Chapter 13 is devoted to extreme temperature measurements both low and high; the methods of attaining them are also discussed. This chapter is followed by one describing a number of of special methods of temperature measurements and a short chapter on the International Temperature Scale.

These fifteen chapters constitute Part I of the book, in which the whole field of temperature measurements and the principles upon which they depend and with which they are associated are treated in a very complete way. In fact, it is comprehensive enough to serve as an excellent textbook for a semester's course in heat for students who have had an introductory course in physics and a course in integral calculus. Each chapter is followed by a goodly number of well-chosen problems.

Part II contains a group of twenty-nine well-chosen laboratory experiments in temperature and heat measurement. They are advanced enough to be stimulating to students of intermediate physics, and yet they are not so complex that they may not be set up from apparatus usually found around a college physics laboratory.

An appendix of fifty pages gives a considerable amount of general information of value to both student and instructor as well as twenty-seven tables of reference matter.

The book is well written, the discussions are clear and explicit. It contains a number of historical references that lend interest to its reading. It contains much information that is not found in the usual textbook on heat; the reviewer recommends it as a reference book for any student of physics.

JOHN G. ALBRIGHT
Rhode Island State College

The Mathematical Theory of Plasticity

BY RODNEY HILL. Pp. 356+ix. Oxford University Press, New York, 1050. Price $7.00.

This excellent book treating the flow of metals will be a standard reference book for the rheologist in the future.

The subject is developed in an orderly fashion. The author, assuming the reader has a working knowledge of stress and strain tensors and the rudiments of elasticity, commences his treatment with the yield conditions and stress-strain relationships involved in plastic flow. He is to be especially commended for the critical and careful, though not belabored, presentation of the latter topic. He develops the plane strain problem very fully and, of course, includes many of the applications of this analysis to technological forming operations (rolling, drawing, extrusion, etc.) which he has published recently in British metallurgical journals. As a matter of fact, this and the phenomenological treatment of plastic flow in the presence of anisotropy occupy most of the latter half of the book. The reader can thus see that the book is specialized in scope: it—not unlike Gatty and Spooner's

recent book of the same publishers—is a coherent and comprehensive review of the author's work of the past five years, preceded by a critical summary of those principles of the science on which his work is based.

The reader will appreciate the obvious care to which the author has gone to refer to primary sources (Huber's Polish paper is a case in point). The system of notation used is brief and useful and is carefully explained in an Appendix. One might, however, wish for a page of nomenclature: this would increase the usefulness of the book for reference, for a reader should hardly be expected to remember the large number of symbols—many of which the author uses in his own manner (σ' for stress deviator, for example).

W. M. Baldwin, Jr.
Case Institute of Technology

Physical Chemistry of High Polymeric Systems, Vol. II

By H. Mark and A. V. Tobolsky. Pp. 506+xi, Figs. XIII-5. Interscience Publishers, Inc., New York, 1950. Price $6.50.

The present volume is a completely revised and augmented edition based on the 1940 American edition by H. Mark, bearing the same title. We plan to review the present volume by comparing it with the 1940 edition, not only because that edition is now very well known to most polymer chemists and physicists in this country, but also because a comparison of the two volumes presents an excellent summary of the major advances which have been made in this important field during the past ten years.

In 1940 the main outlines of the major problems connected with the physical chemistry of high polymeric systems were rather well appreciated by Professor Mark. However, there were not available sufficient data to amplify such theoretical treatments as did exist at that time and to allow one to make any intelligent choice among the various theories. Hence, the 1940 edition was concerned very largely with the physical tools which would be needed to study polymers, such as x-ray diffraction, electron diffraction, dielectric methods; the various molecular weight techniques; solubility, fractionation, and degradation. Most of the numerical examples were chosen either from natural high polymers such as the proteins, or from large organic molecules such as some of the dye molecules. The outline of the 1940 edition was so well conceived that it was able to form the main backbone of the new, revised edition. It was necessary simply to fill in the various gaps with the data which have since been accumulated and to describe some of the newer physical tools for studying polymers and to outline some of the new physical concepts. Among the new physical tools mentioned here are neutron scattering by organic molecules, light scattering from high polymer solutions, and the infrared absorption spectra of numerous polymer systems. Theoretical advances connected with the thermodynamics of polymer solutions, the swelling of polymers, the theory of intrinsic viscosity, and the viscosity of polyelectrolyte solutions have been discussed in a very thorough fashion.

There is an entirely new section on mechanical behavior of high polymers, treating mainly the visco-elastic behavior of polymers, but also extending the kinetic theory of rubber elasticity which had already been presented in the earlier edition. The present volume uses the term "chemorheology" to signify that creep and stress relaxation in polymers frequently depend on complex chemical reactions such as chain scission, cross linking, or even reversible cleavage of chemical bonds as in the polysulfide rubbers.

In the field of reaction kinetics the topics of polymerization, copolymerization, and degradation are thoroughly reviewed. In 1940, the subject of polymerization kinetics had to be treated in very broad and general terms, while copolymerization was not even mentioned. Now specific reaction mechanisms for different monomers under various polymerization conditions are discussed in great detail, and the specific reaction rate constants are tabulated. Copolymerization theory is treated quite fully, and a large table of monomer reactivity ratios is included.

An eight-page section on degradation in the earlier edition has been expanded into a full chapter of 28 pages. New theoretical and experimental results on the degradation of polymer chains have been given. The topics of oxidation, ozonolysis, and other forms of chemical degradation are covered.

In summary, this book is an excellent survey of the present status of the physical chemistry of high polymers. It might well serve as the basis for a graduate level course. The illustrations have been well chosen, while the numerous tables giving compilations of physical constants are extremely helpful.

R. F. Boyer
Dow Chemical Company

An Introduction to Probability Theory and Its Applications, Vol. I

By William Feller. Pp. 419+xii. John Wiley and Sons, Inc., New York, 1950. Price $6.00.

Probability is a difficult subject for a textbook, now that the subject is a mathematical discipline. Up to about 25 years ago, a text on the subject could be essentially a compliation of problems, aside from some semiphilosophical remarks on the nature of probability which the writer would usually have been well advised to omit and the reader to ignore. Now that mathematical probability is understood to be a part of the theory of measure, its logical justification requires no philosophical remarks and the subject can be treated in a logical and systematic way. Unfortunately, such a treatment would make impossible demands on any reader unwilling to go into measure theory rather deeply. There seems to be no reasonable compromise treatment, general enough to cover at least the usual simple problems, yet restricted enough to be understandable to non-measure specialists. Feller's compromise is to deal only with discrete sample spaces. That is, he deals only with cases in which there are finitely or denumerably infinitely many basic events out of which compound event probabilities are then derived. Thus, for example, random variables which have densities of distribution, say, normally distributed random variables, are necessarily omitted.

Feller's is the first modern text on probability. His treatment is logically rigorous without being pedantic, and anyone who finishes the book will have both a fair understanding of the mathematical significance of probability and an acquaintance with a wealth of problems and applications. The limitation to discrete sample spaces makes the book understandable to readers knowing calculus. Besides the usual permutation-combination problems, which are made more than banal formula juggling by their applications, there are valuable treatments of relatively advanced topics. For example, there is a complete and original treatment of Markov chains, based on a deep study of recurrent events. Random walks and birth and death processes are studied, with many applications. Readers of this Journal will find the applications to physics of particular interest.

It is not yet clear how a text covering continuous probabilities can be written, but it is to be hoped that the "I" on the author's title page indicates that he will write it.

J. L. Doob
University of Illinois

Theory of the Interior Ballistics of Guns

By J. Corner. Pp. 443+xiii. John Wiley and Sons, Inc., New York, 1950. Price $8.00.

The theory of interior ballistics deals with the phenomena occurring within a gun during a firing. This book is the first advanced discussion of the subject to appear in English and is the only one to appear for some time in any language. Dr. Corner has himself carried out important research in interior ballistics, and much of the book represents his own contributions to the field.

Several "ballistics systems" are discussed, and examples of applications of each are given in detail. These include a "practical" system, i.e., a rough approximate scheme for predicting the behavior of conventional guns under ordinary conditions. Such a system, most often used in connection with routine proof firings or minor design changes, has the advantages of speed, ease of applicability, and accuracy if suitable empirical constants are available. In addition, the author discusses several approaches to an accurate theoretical system of ballistics. Such a "research" system of ballistics is necessary in the study of unusual guns. The ballistics of a recoil-less gun, i.e., a gun with a nozzle in the breech to prevent recoil, a gun with a tapered bore, and several other unconventional guns are considered.

Various problems which appear in an accurate study of interior ballistics are considered in detail. These include the problem of the nature of the burning of the propellant grains. The author discusses his own contributions to the subject, including a discussion of the propagation of flames in gases. The nature of the flow of the powder gases in the gun barrel behind the projectile and after the ejection of the shot has been the subject of considerable research. This problem is of intrinsic hydrodynamic interest and is of importance in some design problems, such as in the design of muzzle brakes. A chapter is devoted to the problem of the heat lost to the bore during the firing.

The book is well written and is, at present, the only authoritative text on the interior ballistics of guns. It should certainly form the basis of any serious study of the subject.

CHARLES F. CURTISS
University of Wisconsin

The Clinical Use of Radioactive Isotopes

BY BERTRAM V. A. LOW-BEER. Pp. 414+xv, Figs. 58+5 graphs. Charles C. Thomas, Publisher, Springfield, Illinois, 1950. Price $9.50.

The title of this book may be misleading, for it contains not only clinical information but also detailed instruction in radiation physics. Part I, Physics, comprises about one-third of the total volume. In addition, large parts of Part II, Clinical Applications, are devoted to physics, such as, for example, production of radioactive isotopes in high energy accelerators and nuclear reactors, decay curves of radioactive isotopes, dosage calculations. The section on Physics discusses atomic theories, mass, energy and charge relations of atoms, natural and artificial radioactivity, nomenclature of nuclear reactions, and nuclear decay laws. This is followed by chapters on detection and measurement of radioactivity, using radioautographic and other photographic techniques, ionization methods, absorption curves, electrometers, and pulse counters with explanations of possible measuring errors due to the statistical nature of the radioactive decay. Standard units, such as the roentgen, the curie, the rutherford, are presented in addition to more or less temporary units for biological work such as the rep (roentgen-equivalent-physical), the rem (roentgen-equivalent-man), er (equivalent roentgen), rhm (roentgen per hour at one meter), gram-roentgen, energy unit, and J unit. Finally, useful advice is given regarding safeguards against hazards in the use of radioactive isotopes, protection against various types of rays, storage, and disposal of wastes. There are abundant reference lists to books and articles on related subjects and an appendix with numerous definitions and tables. The book is well written and illustrated, and its information of greatest value for anyone using radioactive isotopes for clinical or any experimental use.

OTTO GLASSER
Cleveland Clinic Foundation

Wave Theory of Aberrations

BY H. H. HOPKINS. Pp. 169+viii. Oxford University Press. New York, 1950. Price $3.00.

This is a textbook of geometrical optics, in spite of the title. A knowledge of elementary calculus and paraxial thick-lens optics is required of the reader. As announced in the preface, this is not intended to be a comprehensive textbook of lens designing, since the author felt that the inclusion of this subject would have led to a book of unwieldy size. Nonetheless, this book contains all of the theoretical equipment commonly used by lens designers, and Chapter VIII even gives a detailed account of the author's favorite computing scheme for tabulating the Seidelian aberrations.

Concerning the intended meaning of the title, we quote from the author's preface: "The aberrations of an optical system can be looked upon either as a deviation of the actual wave-front from an ideal spherical one, or alternatively as a failure of the image-forming pencil of rays to unite in a point. While engaged upon practical lens designing I rapidly came to the conclusion that the former approach was very much to be preferred. . . . Regarded as the geometry of wave-normals, the whole of ray optics may obviously be taken over into wave optics. . . . Where a result has been known in ray optics, I have 'translated' it into wave optics. In particular, the informed reader will see where my results are the analogues of those to be found in the books of H. D. Taylor, and A. E. Conrady." Chapter II is devoted to methods of conversion from the one notation to the other.

The author has successfully steered a reasonable middle course between the difficult rigor of modern mathematics at one extreme and fraudulent oversimplification at the other. His own proof of the optical sine condition (taking spherical aberration into account) compares favorably with that given in most textbooks, since its validity is not restricted to rays lying in a meridian plane, and no gratuitous assumptions are made concerning an effective pupil surface.

The concept of the different aberrations as terms in a series expansion is introduced early in the book (Chapter IV). The aberrations are thus defined in the most logical and straightforward manner. Compared with the two acknowledged sources of inspiration, the present text constitutes a commendable condensation of the subject matter into a volume of readable size.

The author's partiality to wave front aberrations will perhaps not be shared by some readers. The advantages of this concept are, of course, the great simplification in the theory of aspheric surfaces (Chapter XI) and the fact that all aberrations are expressed directly in terms of Rayleigh limits. On the other hand, an important disadvantage is the ambiguity pointed out by the author himself on pp. 142–144. In eliminating this, he introduces the awkward "shift of focus term." Also, the derivation of the Seidelian aberrations (Chapter VII) is unnecessarily difficult compared with that given by Conrady.

The theory of orthogonal systems is given in Chapters V and VI without reference to the previous work of Gullstrand, Boegehold, and Herzberger. Chapter VI has no obvious bearing on the rest of the book.

L. IVAN EPSTEIN
Bausch & Lomb Optical Company

Survey of Modern Electronics

BY PAUL G. ANDRES. Pp. 522+x, Figs. 11.28. John Wiley and Sons, Inc., New York, 1950. Price $5.75.

This book was written as a text for a short survey course for students in electrical engineering. The purpose of the course is to give the student an idea of the methods and scope of electronics so that he may decide whether he wishes to specialize in electronics or in some other branch of electrical engineering. The author does not mention whether laboratory experiments are performed by the students, although such experiments are especially effective in stirring up the students' enthusiasm and interest in electronics and in demonstrating the operation of some of the circuits. Perhaps the difficulty of correlating laboratory work with a survey course is too great.

Students should enjoy studying this impressive assemblage of applications of electronics to physics, chemistry, biology, engineering, etc. They should find the book easy to read, since the

treatment is mostly descriptive rather than mathematical. The numerous halftones of actual equipment add greatly to the attractiveness of the text. The idea of following the description of most tubes by practical applications is a good one for maintaining the interest of the students. A few problems and numerous references to other publications are included at the end of each chapter. As Professor Andres runs the gamut from amplification to zipperlength control, some of his better descriptions include electrostatic spraying systems, automatic pilot, x-ray tubes and applications, inverters, direct-reading spectrometer, counters, timers, strain gauges, communication circuits, induction heating, dielectric heating, and welding controls.

These same features which make this book appealing to students also should make it teachable from the point of view of the instructors of survey courses. Such instructors should approve of this book. The opinions of other engineers or physicists will largely depend upon their opinions of survey courses.

It is probably easy to make complaints about a book which discusses several fields of science, but the reviewer would like to note a few. Most of the explanations are clear, with only occasional lapses into illogical grammatical constructions which plague all of us. Some misstatements were noted, but their correction should not be difficult. Many of the complicated examples are discussed too soon after a principle or tube is explained. For example, the automatic pilot, as the main example of amplifiers, is rather difficult to understand unless some simpler control circuits are studied first. Physicists at least might suggest that a proper statement of Ohm's Law would facilitate the explanation of the difference between the resistance of metal conductors and the "resistance" of some semiconductors and electron tubes.

It would perhaps be captious to make a list of omissions in a book which is filled with sufficient material for its intended purpose. It was disappointing, however, not to find a circuit for the common voltage-regulated power supplies or a careful discussion of the feed-back principle. Also, there is nothing about magnetic amplifiers, although several circuits do employ saturable reactors. In the other direction, it would seem that nuclear fission and the atomic bomb are topics which are more suitable for other courses.

J. McCarthy
Western Reserve University

Photography in Astronomy

By E. W. H. Selwyn. Pp. 112, Figs. 48. Eastman Kodak Company, Rochester, New York, 1950. Price $2.75.

The importance of photography as the technique of most general application in astronomy requires no elaboration, and any book on the subject from the Kodak Research Laboratories is sure to be welcomed. In the present volume Dr. Selwyn has skillfully condensed into one-hundred-odd pages an excellent description of the characteristics of photographic materials, the methods of using them, and the pitfalls which lie in the paths of those who apply such methods.

The first two chapters describe typical subjects of astronomical photography and the telescopes and auxiliary equipment employed. These chapters are necessarily elementary and will not be of great value to readers who are equipped to profit by the remainder of the book.

Chapters 3, 4, and 5 comprise about half of the volume, and contain the matter of greatest astronomical importance under the headings "Properties of photographic materials," "Fine structure of photographic images," and "Photometry." As would be anticipated in the case of any publication from this source, these chapters lack nothing in accuracy and, within space limitations, draw attention to many major problems and techniques for their solution. The discussions of sensitometry, characteristic curves, and reciprocity failure, together with interpretation in terms of latent image theory, are the most detailed and valuable portions of the book. By comparison, the sections on stellar photometry and spectrophotometry are so short that little hint is given of the complexities encountered in practice. Astronomers will note the omission of reference to the astrometrically important type of emulsion creep discovered some years ago by K. Aa. Strand.

Chapter 6 on "Processing" surveys essentially the same ground that is covered by the inexpensive pamphlets previously issued by Eastman Kodak Company. Chapter 7, "Reproduction," offers useful suggestions for enhancement of contrast and the reproduction and normalization of stellar spectra.

The illustrations include a number of excellent examples of astronomical photography, among which extragalactic nebulae are conspicuously absent.

A few errors of astronomical fact were noted. Figure 8 includes an energy distribution curve for spectral class *L*, which was discarded as a representative stellar type by 1900. On page 17 it is stated that cool stars are intrinsically fainter than hot stars, a proposition which ignores the dependence of stellar luminosity on radius.

Omissions of letters or figures will be found in Figs. 28 and 32a, and misprints in the second equation on page 57 and the second on page 83. The final equation on page 83 is correct only in the special case of an AO star of apparent magnitude zero.

This book is by no means intended to be a monograph or text on astronomical photography; its very conciseness will invite the student to become familiar with the elements of the subject, and it should certainly be ready to the hand of all who are making a start in astronomical research.

Freeman D. Miller
Observatory, University of Michigan

Electronenoptik, Band I—Grundzüge der theoretischen Elektronenoptik

By Alexander A. Rusterholz. Pp. 249, Figs. 118. Verlag. Birkhäuser, Basel, 1950.

The present volume, which is to be followed by a second one covering applications, is a methodical presentation of the principles of electron optics by a member of the staff of the tube plant of Brown, Boveri and Cie in Baden (Switzerland). After establishing the basis for electron ray optics with the aid of the lagrange equations, the author concentrates his attention on the lens properties of axially symmetric fields. In logical sequence, he treats the calculation and measurement of the fields, the determination of the electron paths, the cardinal points of electric and magnetic lenses, and lens aberrations. Two-dimensional fields, deflecting fields, and electron mirrors are covered in less detail.

The book is written in textbook style and presents its material in unhurried fashion, giving its mathematical derivations in rather full detail. Emphasis is placed on the most commonly encountered problems of electron optics. The author has succeeded well in giving a unified presentation of the pertinent literature and, through well-placed references, spurs the reader on to delve further on his own. There are only a very few errors; and these are, as far as the reviewer has been able to note, of a rather trivial nature. In brief, the book can be recommended to the student as a well-balanced introduction to the theory of electron optics (of static fields without space charge) and to the research worker as a convenient repository of the principal formulas of the subject.

E. G. Ramberg
Radio Corporation of America

The Properties of Metallic Materials at Low Temperatures

By P. Litherland Teed. Pp. 222+viii. John Wiley and Sons, Inc., New York, 1950. Price $3.50.

This booklet is primarily a compilation of data on the effects of low temperatures on the yield strength, tensile strength, elongation, reduction of area, and impact energy of the common metals and alloys. These comprise approximately one-half the rather small total volume of the book, and this leaves some hundred

pages for the discussion of the subject. The space is also nearly evenly divided between ferrous and nonferrous alloys, and particular attention has been paid to the effects of welding on ferrous materials. The author succeeds in presenting the major problems encountered in low temperature applications in a lucid and interesting manner. The booklet will be, therefore, useful to everybody who desires to become rapidly acquainted with the general problems and who requires some general information on the low temperature properties of various metals and their alloys, in general, and of various types of ferrous alloys, in particular. On the other hand, the author's discussions deal mostly with practical observations, but only in a very condensed manner with generalizations, and still less with the many baffling fundamental problems encountered at low temperatures. The extensive bibliography appended to each chapter of the booklet, therefore, will be useful to those who are more deeply concerned with the subject matter of the booklet reviewed.

R. G. SACHS
Horizons, Inc.

Industrial High Frequency Electric Power

BY E. MAY. Pp. 355+xi, Figs. 9–34. John Wiley and Sons, Inc., New York, 1950. Price $5.00.

The book deals with induction heating and dielectric heating. The presentation is on graduate level. In the reviewer's opinion, it is the first book which covers induction and dielectric heating with gratifying breadth and depth, necessary for the full appreciation of the problems involved. The author is congratulated on the superb accomplishment he has achieved in writing this book, which, undoubtedly, will assume a most prominent position in the field of high frequency heating.

The book begins with a summary of basic circuit theory, with special emphasis on tuned circuits. Proper attention is given to air-cored auto transformers which are widely used for high current inductors.

A description of the various types of high frequency generators follows. First, the spark-gap oscillator is described, and numerical data regarding its circuit parameters is presented. Then a discussion follows on mercury-arc inverters, both single-phase and three-phase. Next, a very interesting chapter on rotating high frequency generators follows. Finally, the vacuum tube oscillators are analyzed with resistive and tuned load. A great wealth of information, both theoretical and practical, is incorporated in this chapter.

Subsequently, the theory of induction heating is treated. The heating of solid cylindrical charges, hollow cylindrical charges, and flat strips is analyzed. Expressions for the calculation of inductors are given. Surface hardening is numerically treated, and many applications are described.

An analysis of the fundamentals of dielectric heating follows. Power oscillators in the megacycle range are described, and the problems of transmitting megacycle power are discussed. Auxiliary equipment to the power oscillator such as electronic rectifiers, timing devices, and safety measures are discussed. Attention is also placed on screening of high frequency equipment. Techniques for the measurement of high frequency voltage, current, power, etc., conclude this chapter. The last chapter deals with industrial applications and operating problems and is—like the rest of the book—very well illustrated.

The book deserves the highest degree of recommendation.

H. F. STROM
General Electric Company

Colloidal Dispersions

BY EARL K. FISCHER. Pp. 387+vii. John Wiley and Sons, Inc., New York, 1950. Price $7.50.

The author, now a technologist with the National Bureau of Standards, is a research chemist with several years' experience in problems relating to paint and ink manufacture and to textile technology.

The scope of the book is limited to dispersions of solids in liquids and, among those, mainly to the dispersions of pigments. The method of preparation usually is by grinding of the dry solid and subsequent mixing with the dispersion medium. Several chapters therefore are given to the description of industrial mills and mixers, and an additional chapter to the flushing process (direct transfer of the dispersed pigment from the aqueous to the oil phase). Dispersion by ultrasonic waves is briefly discussed.

The mixtures obtained by the above methods sometimes do not reach the colloid state of subdivision (see pp. 54–67 regarding the particle sizes). Sedimentation on standing and abnormal flow are often observed. The chapters dealing with the physical properties of the dispersions, which make up the greater part of the book, have more to offer to a reader with a general interest in colloid science. The discussion of particle size measurement, of surface forces and wetting, and of the rheological properties is especially thorough and comprehensive. The chemical aspects of colloid behavior are less extensively presented, and a physicist wishing to use the book will not encounter a chemical equation until page 253, in the chapter on surface active agents.

The text is copiously documented with references to recent periodical and patent literature. It should be of great value to scientists, chemical engineers, and others engaged in the development and manufacture of paints and printing inks. The presentation is from a broad viewpoint, with emphasis on the underlying principles. Much of the information should therefore find use in other fields of industrial research, wherever similar dispersions are prepared and studied.

A few errors probably are unavoidable in a first edition. The reviewer has noticed one on page 15, line 6 from top, and two on page 253, line 15 from top.

BASIL C. SOYENKOFF
New York University

The Acceleration of Particles to High Energies

BY THE INSTITUTE OF PHYSICS (LONDON). Pp. 58+x, Figs. 4.5. Institute of Physics, London, 1950.

This book is based on certain papers presented by the Electronics group at the Institute of Physics at a convention in May, 1949.

In his foreword, Professor M. L. E. Oliphant points out the necessity in nuclear physics of machines that will produce high energy particles in the range from 10 million to one billion electron volts, the large power consumption, and the great expense of such apparatus.

In the first chapter the fundamental principles of the cyclotron and the synchrocyclotron and their general construction are explained and described. Some of the results obtained in America and details of British synchrocyclotrons under construction are given. A selected bibliography of 18 items concludes this chapter. No reference is made to the "electron cyclotron" developed by Henderson at the Canadian National Research Council.

The second chapter covers betatrons and synchrotrons. Nuclear research work such as photodisintegration and gamma-neutron reactions is mentioned, and the medical application of 30-Mev x-rays in cancer therapy is called a promising possibility. In connection with the so-called depth dose effect of high energy x-rays, the author cites a German reference of 1943 instead of the work by Charlton and Breed of 1948; consequently, his example is incorrect. A fairly complete bibliography concludes the chapter. The 1937 Steenbeck "betatron" patent, the depth dose article mentioned above, and a reference to the description of the first American synchrotron might have been included.

A short third chapter covers electrostatic generators. In the discussion of insulating gases "silicon fluoride" was named where sulfur hexafluoride was meant. The fire hazard of compressed air and the uses of nitrogen and Freon are discussed. In a paragraph on "The Accelerating Tube" the failure of longer tubes at lower gradients is mentioned. A bibliography concludes the chapter.

The fourth chapter is concerned with linear accelerators, a field in which the British have done very successful pioneer work. Wave guide accelerators of the traveling wave type for electrons are described and illustrated, in particular D. W. Fry's 4-Mev accelerator, in which an average current of 60 microamperes was obtained. A similar machine for 10 Mev is under construction. Conservation of rf power by feedback from the end of the wave guide to the start and the use of ceramic loading diaphragms are mentioned as future developments. A short bibliography concludes the chapter. No reference is made to the work of the late W. W. Hanson and his associates at Stanford University.

A final chapter is formed by the discussion which followed the papers at the original conference and additions made by correspondence.

The book offers convenient nonmathematical descriptions and explanations of the various circular and linear particle accelerators that have become important in nuclear research, industry, and medicine. It will be very useful to those physicists, engineers, and medical workers who want a bird's-eye view of the whole field and do not wish to go into technical details. There are 9 excellent photographs of various machines and 15 clearly drawn diagrams. A subject index and a name index are added at the end of the book.

W. F. WESTENDORP
General Electric Research Laboratories

Other Books Received

An Index of Nomograms. D. P. ADAMS. Pp. 174. The Technology Press and John Wiley and Sons, Inc., New York, 1950. Price $4.00.

Theory and Application of High Frequency Phenomena. Y. ASAMI. Pp. 68+iii, Figs. 38. Research Institute of Applied Electricity, Hokkaido University, Sapporo, Japan, 1950.

TV and Other Receiving Antennas. ARNOLD B. BAILEY. Pp. 595. Figs. 310. John F. Rider, Publisher, Inc., New York, 1950. Price $6.00.

Vorstufe zur Theoretischen Physik. RICHARD BECKER. Pp. 172, Figs. 94. Springer-Verlag, Berlin, 1950.

Voltage Stabilizers. F. A. BENSON. Pp. 125, Figs. 97. Electronic Engineering, Strand, London, England, 1950.

Nuclear Physics. FRANCIS BITTER. Pp. 200+viii. Addison-Wesley Press, Inc., Cambridge, Massachusetts, 1950. Price $5.50.

Unit Operations. GEORGE GRANGER BROWN. Pp. 611+xi, Figs. 551. John Wiley and Sons, Inc., New York, 1950. Price $7.50.

Radioactivity and Nuclear Physics. JAMES M. CORK. Pp. 415+xiii. D. Van Nostrand Company, Inc., New York, 1950. $5.00.

The Special Theory of Relativity. HERBERT DINGLE. Pp. 94+vii, Figs. 2. Methuen's Monograph No. 2, John Wiley and Sons, Inc., New York, 1950. Price $1.25.

Sourcebook on Atomic Energy. SAMUEL GLASSTONE. Pp. 546. D. Van Nostrand Company, Inc., New York, 1950. Price $2.90.

Kinematics of Machines. GEORGE L. GUILLET AND A. H. CHURCH. Fifth Edition. Pp. 299+vii. John Wiley and Sons, Inc., New York, 1950. Price $4.00.

Electromagnetic Theory. OLIVER HEAVISIDE. Pp. 386. Dover Publications, Inc., New York, 1950. Price $7.50.

Kleinste Drucke ihre Messung und Erzeugung. RUDOLF JAECKEL. Pp. 302, Figs. 301. Springer-Verlag, Berlin, 1950.

Fundamentals of Optics. FRANCIS A. JENKINS AND HARVEY E. WHITE. Pp. 647+x. McGraw-Hill Book Company, Inc., New York, 1950. Price $7.00.

Die Electromagnetische Schirmung in der Fernmelde- und Hochfrequenztechnik. HEINRICH KADEN. Pp. 274, Figs. 145. Springer-Verlag, Berlin, 1950.

Theory of Sets. E. KAMKE (translated from the German by F. Bagemihl). Pp. 144, Figs. 4. Dover Publications, New York, 1950. Price $2.45.

Physics, Its Laws, Ideas and Methods. ALEXANDER KOLIN. Pp. 890+xvi. McGraw-Hill Book Company, Inc., New York, 1950. Price $6.50.

Wave Guides. H. R. L. LAMONT. Pp. 118+vii, Figs. 37. Methuen's Monograph No. 1, John Wiley and Sons, Inc., New York, 1950. Price $0.50.

Die Mathematischen Hilfsmittel des Physikers. E. MADELUNG. Fourth Edition, enlarged and improved. Pp. 531, Figs. 29. Springer-Verlag, Berlin, 1950.

TV Installation Techniques. SAMUEL L. MARSHALL. Pp. 330. John F. Rider, Publisher, Inc., New York, 1950. Price $3.60.

The Atomic-Molecular Theory. LEONARD K. NASH. Pp. 115. Harvard University Press, Cambridge, Massachusetts, 1950. Price $1.35.

Nuclear Data. National Bureau of Standards Nuclear Data Group: KATHARINE WAY, LILLA FANO, MILLICENT R. SCOTT, AND KARIN THEW. Pp. 309+xiv. Superintendent of Documents, Washington, 1950. Price $4.25.

Table of the Bessel Functions $Y_0(Z)$ and $Y_1(Z)$ for Complex Arguments. The Staff of National Bureau of Standards Computation Laboratory. Pp. 427, Figs. 4. Columbia University Press, New York, 1950. Price $7.50.

The Identification of Molecular Spectra. R. W. B. PEARSE AND A. G. GAYDON. Second Edition. Pp. 276+xi. John Wiley and Sons, Inc., New York, 1950. Price $8.50.

Electrical Engineers Handbook, Vol. II. HAROLD PENDER AND KNOX MCILWAIN. Fourth Edition. John Wiley and Sons, Inc., New York, 1950. Price $8.50.

Fundamentals of Quantum Mechanics. ENRICO PERSICO (translated by Georges M. Temmer). Pp. 484+xii, Figs. 48. Prentice-Hall, Inc., New York, 1950. Price $6.00.

Mathematics of Relativity. G. Y. RAINICH. Pp. 173+vi. John Wiley and Sons, Inc., New York, 1950. Price $3.50.

Leitfaden der Technischen Wärmelehre nebst Anwendungsbeispielen. H. RICHTER. Pp. 617, Figs. 384. Springer-Verlag, Berlin, 1950.

Encyclopedia on Cathode-Ray Oscilloscopes and Their Uses. JOHN F. RIDER AND SEYMOUR D. USLAN. Pp. 982+vii. John F. Rider, Publisher, Inc., New York, 1950. Price $9.00.